Strategies for characterisation of the unexplored human phosphoproteome

A thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy

by

Gemma Elizabeth Hardman

September 2017

Acknowledgments

Firstly, I would like to thank my supervisor, Professor Claire Eyers. I am very grateful to have had the opportunity to learn so much about mass spectrometry, phosphoproteomics and biochemistry, and for all the support I have received along the way.

Thank you to Dr Pat Eyers for help and enthusiasm about all things related to kinases and to Dr Dada Pisconti for assistance with cell culture. Thank you to Professor Natarajan Kannan and Zheng Ruan for the molecular dynamics simulations, and to Professor Andy Jones and Simon Perkins for (a lot of) assistance with the bioinformatics. Also thank you to Dom Bryne and Fiona Bailey for helping me with many new lab techniques.

Thank you to Professor Rob Beynon and everyone in the Centre for Proteome Research; it has been a pleasure to work with you all. I am very grateful for your support, friendship and all the knowledge I have gained working alongside you. A special thank you is reserved for Stephen Holman and Philip Brownridge who, despite calling me 'student' for the first year, have provided endless help in all things proteomics and mass spectrometry related.

A huge thank you to Becka, Grace and Sam for the cheese & wine nights and the 'civilised' meals, for the many hours spent on the dancefloor of Brooklyn Mixer, and for always being there for a cup of tea and a chat. I couldn't have done it without you.

To all my friends, near and far, thank you for always asking 'How's the PhD going?', even if you still haven't got a clue what it's about!

Thank you also to my family; I'm so grateful for all your love and support, and I hope I continue to make you proud. A special thanks goes to Zoë (aka Bobby) for sharing pasta bakes with me, always making me laugh, and just generally being the best sister I could wish for.

And finally, to Jonny: your support these last four years has been invaluable, and in the final few months I'm especially grateful to have shared many cups of tea, glasses of wine and the Friends boxset with you! Thank you for everything.

Abstract

Title: Strategies for characterisation of the unexplored human phosphoproteome

Author: Gemma Elizabeth Hardman

Post-translational modification of proteins by addition of a phosphate group serves as a means of rapidly regulating protein function, and thus has vital roles in all aspects of cell biology. Phosphorylation of serine, threonine and tyrosine are the focus of the vast majority of studies aimed at elucidating the function of such modification in human cells. However, there is growing evidence that phosphohistidine (pHis) may also have a role in diverse cellular processes in mammalian systems.

Characterisation of pHis is compromised by the inherent instability of the acid-labile phosphoramidate bond, meaning there is a severe lack of suitable experimental strategies to pinpoint the presence of pHis in proteins. The acidic conditions integral to current phosphoproteomics workflows are incompatible with analysis of acid-labile pHis.

In this work, chemical phosphorylation was used to generate a set of model pHis-containing peptides, for which stability of the histidine phosphorylation across a range of pH values and at elevated temperature was assessed. The pHis-containing peptides were further used to evaluate various existing enrichment strategies for their suitability for pHis. None of the enrichment methods tested for this work that had previously been used for enrichment of serine/threonine phosphopeptides enabled recovery of pHis peptides.

Subsequently, a strong anion exchange fractionation method has been developed, which permits separation of pHis-containing peptides from non-phosphorylated equivalents with salt gradient elution at non-acidic pH. Fractions at the end of the gradient are significantly enriched for phosphopeptides. Crucially, the unbiased phosphopeptide enrichment by strong anion exchange (UPAX) method enabled recovery of pHis-peptides from complex cell lysates.

The UPAX method with characterisation by LC-MS/MS has revealed extensive phosphorylation of histidine, lysine, arginine, aspartate and glutamate in human cell extract, including 310 pHis sites and over 1000 sites of lysine phosphorylation. Remarkably, the extent of non-canonical phosphorylation far exceeds that of phosphotyrosine, exposing the previously underappreciated diversity of the human phosphoproteome.

The results of this work highlight possible roles for pHis and other sites of non-canonical phosphorylation in human cells. Identification of pHis at a conserved residue within the catalytic domain of eukaryotic protein kinases, suggests histidine phosphorylation may act as a general mechanism for kinase regulation. Application of the UPAX strategy for unbiased phosphoproteomics analysis has opened up diverse avenues for exploring roles and regulation of non-canonical phosphorylation sites in any proteome.

Table of Contents

List of Figures

List of Tables

Tables 9.1 – 9.11 are provided in electronic format (details in Appendix)

Abbreviations

ACN	Acetonitrile
AGC	Automatic gain control
AmBic	Ammonium Bicarbonate
ATP	Adenosine triphosphate
BPI	Base peak ion chromatogram
CI	Chemical ionisation
CID	Collision induced dissociation
DDA	Data dependent acquisition
DEPC	Diethylpyrocarbonate
DIA	Data independent acquisition
DMEM	Dulbecco's modified Eagle's medium
DTT	Dithiothreitol
ERLIC	Electrostatic repulsion hydrophilic interaction chromatography
ESI	Electrospray ionisation
ETD	Electron transfer dissociation
FDR	False discovery rate
FLR	False localisation rate
GO	Gene ontology
HAP	Hydroxyapatite
HCD	Higher-energy collisional dissociation
HILIC	Hydrophilic interaction chromatography
HNE	4-hydroxynonenal
IMAC	Immobilized metal ion affinity chromatography
IP	Immunoprecipitation
iTRAQ	Isotope tags for relative and absolute quantification
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
MALDI	Matrix-assisted laser desorption ionisation
MOAC	Metal oxide affinity chromatography
MS	Mass spectrometry
MSA	Multistage activation
m/z	Mass to charge ratio

NT	Non-targeting
PBS	Phosphate buffered saline
PD	Proteome Discoverer
PSM	Peptide spectrum match
PTM	Post-translational modification
PHPT1	Phosphohistidine phosphatase 1
RF	Radio frequency
RP	Reversed phase
SAX	Strong anion exchange
SCX	Strong cation exchange
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SILAC	Stable isotope labelling by amino acids in cell culture
siRNA	Small interfering ribonucleic acid
TFA	Trifluoracetic acid
TMT	Tandem mass tags
XIC	Extracted ion chromatogram

Amino Acids

Amino acid	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

**pHis, pSer, etc = phospho-amino acid

1. Introduction

1.1. Protein Phosphorylation

Protein phosphorylation is a vital post-translational modification (PTM) for the control of cellular processes and adaptation of cells to their environment. Reversible protein phosphorylation is involved in many cellular processes including carbohydrate and lipid metabolism, DNA transcription and replication, memory, cell migration, muscle contraction and cell proliferation and differentiation[1]. Phosphorylation of proteins was first identified in 1906[2], although it was not until 1954 that the first enzymatic phosphorylation was discovered[3]. Since then it has been estimated that one-third of all proteins in eukaryotic cells are phosphorylated at any given time[4]. The most widely studied phosphorylated amino acids are serine (Ser), threonine (Thr) and tyrosine (Tyr), the structures of which are shown in Figure 1.1.

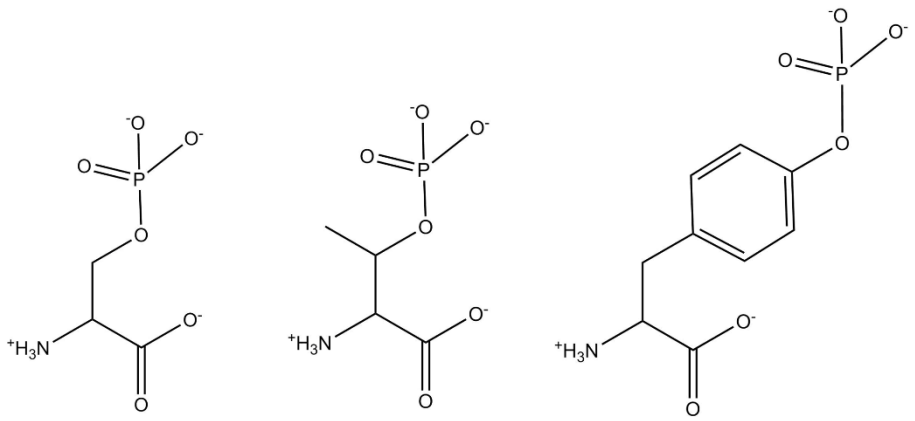


Figure 1.1. Structures of phosphoserine, phosphothreonine and phosphotyrosine

Reversible protein phosphorylation is regulated by kinases and phosphatases. Kinases enable the transfer of a phosphate group from adenosine triphosphate (ATP) to a specific amino acid in a substrate protein, whilst phosphatases are responsible for the removal of phosphate groups from proteins. The addition of a phosphate group to a protein can cause significant conformational changes, thus changing the activity of the protein. Changes in localised charge, and electrostatic and hydrophobic interactions can also be conferred by the addition (or removal) of a phosphate group. Thus, protein phosphorylation modulates

inter- and intra-cellular signalling, and can also regulate protein stability, protein binding and sub-cellular localisation. The site and stoichiometry at which phosphorylation occurs has a direct impact on protein function, therefore identifying sites of phosphorylation is vital for understanding how complex networks of phosphorylation events modulate cellular activity. Abnormal phosphorylation is implicated in a variety of diseases, such as cancer, diabetes and rheumatoid arthritis, further highlighting the importance of studying protein phosphorylation in human systems.

As well as Ser, Thr and Tyr a number of other amino acids can be phosphorylated, including histidine (His), arginine (Arg), lysine (Lys), aspartate (Asp), glutamate (Glu) and cysteine (Cys)[5]. Whilst there are reports of these modifications and/or related kinase activity occurring in mammalian systems, the extent of this non-canonical phosphorylation on human proteins and the possible functions they may have are still largely unknown[6-8]. The characterisation of pHis in human cells is the primary focus of this work.

1.1.1 Phosphohistidine

Histidine can be phosphorylated in either the 1- or 3- positions, resulting in two biologically relevant pHis isomers (Figure 1.2).

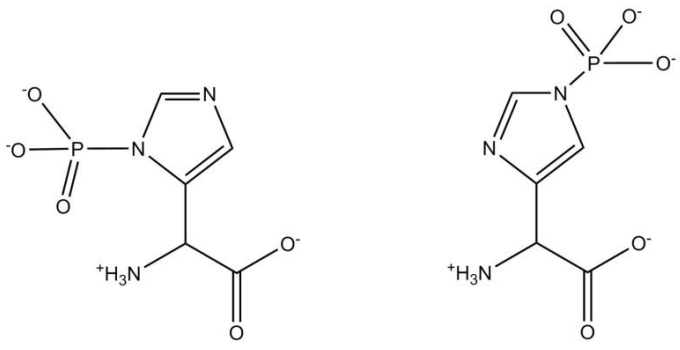


Figure 1.2 Structures of 1- and 3-phosphohistidine

The function of pHis in prokaryotes, fungi and plants is well documented in the context of the two-component system of intracellular signalling, but there is no evidence for two-component signalling in mammalian systems. It has however been estimated that histidine phosphorylation accounts for ~6% of the total phosphorylation in nuclear extracts of model eukaryotes[9], meaning this modification may be more extensive than pTyr, although the

abundance of pHis in the cytoplasm is yet to be assessed. There is little doubt about the presence of histidine phosphorylation in mammalian systems which was first reported in human cells over 50 years ago[10], and there is growing evidence for the roles pHis may have in cellular signalling in these systems. However, widespread characterisation of histidine phosphorylation and its functions in human cells is still lacking.

Two-component systems

The two-component system is a widespread and fundamental mechanism of signal transduction in prokaryotes. Based on the phosphotransfer reaction between His and Asp residues, the two-component system is responsible for conversion of extracellular signal recognition to a cellular response, i.e. via transcriptional regulation. Although eukaryotic signalling is dominated by Ser, Thr and Tyr phosphorylation, the two-component system has been reported in a number of lower eukaryotes including yeast, fungi and plants[11, 12].

The two-component system consists of a sensor histidine kinase and a response regulator (Figure 1.3). Functional histidine kinases in the two-component system are homodimeric. When the sensor domain of the histidine kinase receives an appropriate stimulus (e.g. ligand binding), a conformational change occurs. This brings the kinase domain of one subunit close to the His phosphorylation site of the other subunit, enabling phosphorylation on this His residue to occur. Phosphotransfer from a His to an Asp residue located in the receiver domain of the response regulator then ensues. This Asp phosphorylation event relieves inhibition of the output domain which enables the downstream function, usually transcriptional activation, to be performed. In eukaryotic systems sensor histidine kinases are almost exclusively hybrid kinases, whereby the histidine kinase also contains a response regulator domain.

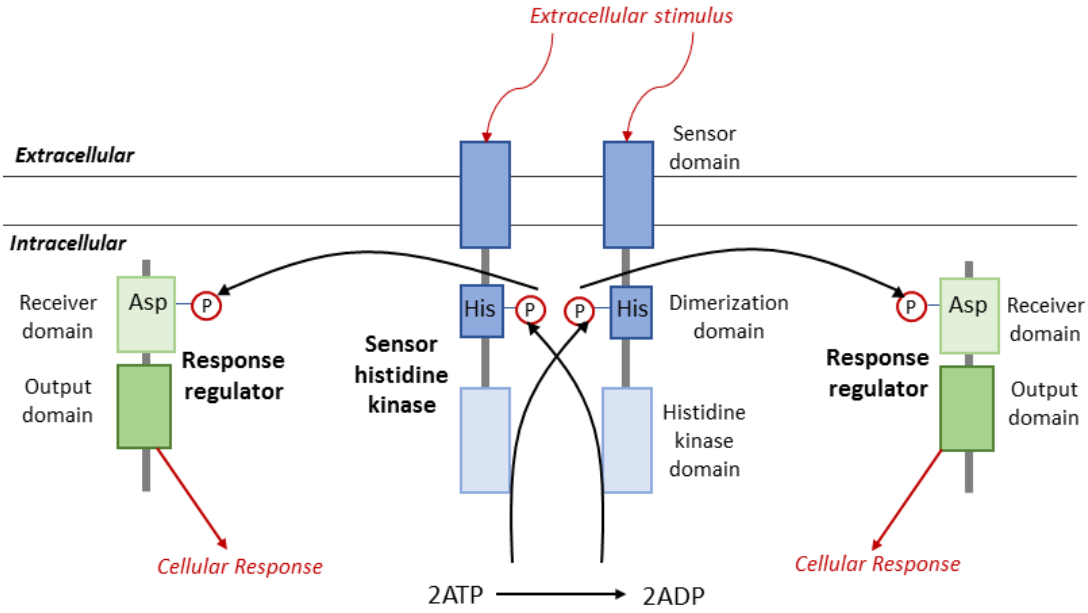


Figure 1.3 Schematic of the two-component system. Phosphorylation of a His in the dimerization domain of a sensor histidine kinase by the histidine kinase domain of the other subunit in the homodimer occurs in response to binding of an external stimulus to the sensor domain. This phosphoryl group is transferred to an Asp residue in the receiver domain of the response regulator which initiates a cellular response (adapted from Besant *et al.*[13])

The typical two-component system is exemplified by the osmoregulatory system in *E. coli*. The mechanism of His-Asp phosphotransfer between the sensor histidine kinase EnvZ and response regulator OmpR enables detection of extracellular osmotic conditions and a resulting transcriptional response to regulate expression of outer membrane proteins OmpF and OmpC[14]. In eukaryotes two-component systems typically consist of more than two proteins. A His-containing phosphotransfer factor can acquire and transfer a phosphoryl group from and to a receiver domain to form a sequential His-Asp-His-Asp phosphotransfer pathway, known as a "multistep phosphorelay". For example, in the fission yeast *S. pombe*, phosphorelay signalling between three hybrid histidine kinases (Phk1/Mak2, Phk2/Mak3 and Phk3/Mak1), a His-containing phosphotransfer (Spy1) and two downstream response regulators (Mcs4 and Prr1) are implicated in oxidative stress responses, mitotic control and sexual development[15]. This type of phosphorelay mechanism also extends to signalling in plants. The first histidine kinase identified in plants was an ethylene sensor ETR1, discovered in *Arabidopsis thaliana*. Further identification of histidine kinases, His-containing phosphotransfer factors and responses regulators indicate a role for the two-component system in ethylene and cytokinin signalling. These plant-

specific hormones regulate almost all aspects of a plants life cycle, including seed germination, fruit ripening and flower development[16].

pHis in mammalian systems

In mammalian systems, transmembrane Tyr kinase receptors initiate similar responses to external stimuli that are performed by two-component systems in prokaryotes and lower eukaryotes. Sequence comparisons indicate the existence of a few proteins in higher eukaryotes with His kinase or response regulator motifs analogous to those in two-component systems, but there is no strong evidence for their wide-spread existence or functional role in mammalian systems[17]. The presence of pHis in mammalian systems has however been known for some time, with growing evidence for its involvement in mammalian signalling. Phosphorylated His residues are known to function as catalytic intermediates in phosphotransfer reactions[18-20], an example being phosphorylation of a conserved His residue within the catalytic core of the histidine tyrosine phosphatase, TULA-2. Thus, pHis has been implicated in glycoprotein VI signalling in platelets and bone remodelling[21, 22]. His phosphorylation also functions as a mechanism for reversible regulation of protein function. Phosphorylation of His has been implicated in the differentiation of PC12 neuronal cells, with nerve growth factor inducing a 3-fold increase in levels of histidine kinase activity against an exogenous substrate[23]. Moreover, Ras-dependent His phosphorylation of a 38 kDa protein from rat liver plasma is differentially regulated during cell division[24]. The kinases and phosphatases involved in pHis regulation are however still largely unknown.

The only mammalian histidine kinases reported to data are NME1 and NME2 (also known as nucleoside diphosphate kinases A and B; NDPK-A and NDPK-B), both of which are implicated in cancer and tumour metastasis[25-27]. With respect to histidine kinase activity a handful of NME1/2 substrates are currently known, including the β-subunit of heterotrimeric G-proteins, the metabolic enzyme adenosine 5-triphosphate-citrate lyase (ACL), the Ca^{2+} activated K^{+} channel KCa3.1 and the epithelial Ca^{2+} channel TRPV5[28-31]. Regulation of His phosphorylation of these proteins is further achieved by activity of phosphohistidine phosphatase 1 (PHPT1). This is an evolutionarily conserved 14 kDa protein which until very recently was the only known example of a His phosphatase in mammalian systems[32, 33]. PHPT1 has been shown to directly bind to and inhibit KCa3.1 by dephosphorylating His358, a His residue which is phosphorylated by NME2[34], thus

indicating a role for pHis in T-cell signalling. Phosphoglycerate mutase 5 (PGAM5) has also been shown to demonstrate His phosphatase activity, specifically against NME2, resulting in inhibition of NME2 mediated His phosphorylation of KCa3.1[35]. Interestingly, it seems as though PHPT1 and PGAM5 dephosphorylate different isomers of pHis, at least in this example.

A number of other proteins have been identified as containing pHis, including P-selectin[36], annexin-I[37] and histone H4[38], although the kinases responsible for their phosphorylation are yet to be discovered. In the case of histone H4, kinase activity was first reported in nuclei of rat liver cells and Walker-256 carcinoma cells in the 1970s[39, 40], but despite similar activity being observed in a number of tissues and cell types, purification and identification of the kinase has not yet been achieved. Histone H4 kinase activity appears to be associated with proliferating cells; there is a large increase in activity preceding proliferation of hepatocytes, and high histone H4 kinase activity is observed in hepatocellular carcinoma cells whilst activity in surrounding tissues is low[41-43].

The slow progress in studying pHis, particularly on a proteome wide scale, stems from the technical difficulties associated with analysis of this labile modification. The inherent instability of His phosphorylation under acidic conditions and at elevated temperature means that typical biochemical techniques and proteomics workflows, such as gel electrophoresis and acidic liquid chromatography (LC) conditions, are not necessarily suited to analysis of pHis (as discussed further in *section 1.5*). Consequently, there is still much to be uncovered about the roles of this modification in mammalian systems.

1.2 Proteomics

Proteomics is the study of the complete protein complement expressed by a genome[44]. A proteome is a dynamic population of proteins, defined by a given cell at a specific time, with consideration required for protein isoforms and modifications[45]. The field of proteomics as a whole covers the determination of protein structure, protein-protein interactions, expression levels and localisation of modifications. Techniques for proteomics analysis include array-based systems, structural and imaging tools, and mass spectrometry (MS), with the latter being the focus of this work.

Proteomics was originally focused on the use of 2D gel electrophoresis, however this was not without its limitations, including difficultly analysing hydrophobic and large molecular weight proteins, a limited dynamic range, problems with visualisation methods and issues with reproducibility[46, 47]. Although for a time 2D gel electrophoresis was used alongside MS techniques, more recent advances have seen a move away from the use of 2D gels. Modern MS approaches enable in depth proteomics analysis of a variety of biological systems through the use of liquid chromatography-tandem mass spectrometry (LC-MS/MS) experiments. It is arguably the vast developments in MS that have driven the field of proteomics forwards so rapidly in the last 20 years.

1.2.1 Mass spectrometry

A general schematic of a mass spectrometer is shown in Figure 1.4, consisting of four main elements: ionisation source, mass analyser, detector and a data system. The ionisation source generates ions and introduces them into the mass spectrometer. From here ions are guided to a mass analyser where they are separated based on their mass to charge (*m/z*) ratio. Ions are detected and information recorded to generate a mass spectrum; *m/z* is plotted against the ion intensity. Typically, for proteomics experiments tandem MS analysis will be performed. The first stage (MS1) is the recording of a full mass spectrum, i.e. detection of all ions across a designated *m/z* range at a given time. Ions of a desired *m/z* are then isolated and undergo fragmentation, with the resulting product ions recorded in a second stage of mass analysis (MS2). A number of fragmentation techniques are available for the purpose of generating product ions, such as collision induced dissociation (CID) and electron transfer dissociation (ETD), which are discussed in more detail in *section 1.2.1.3.* Tandem MS analysis can be performed either in time, i.e. all stages of MS analysis are

performed sequentially in the same mass analyser, or in space whereby MS1 and MS2 are conducted in different mass analysers within the mass spectrometer. In a Top*N* approach to tandem MS analysis, the 'n' most abundant precursor ions are selected to undergo MS2 analysis. Typically, a 'dynamic exclusion' time window is set, during which ions with an *m/z* that has already been fragmented will not be selected for fragmentation again.

Many modern mass spectrometers are hybrid instruments, which incorporate various combinations of mass analysers. The Bruker Amazon mass spectrometer used in this work consists of just one 3D ion trap, whilst the Thermo Orbitrap Fusion mass spectrometer incorporates a quadrupole mass filter, linear ion trap and an Orbitrap.

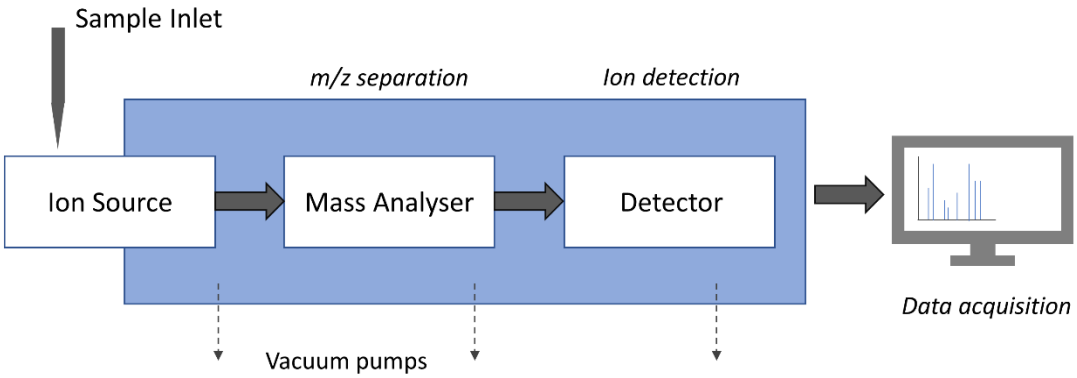


Figure 1.4 Schematic of a typical mass spectrometer. Sample is introduced into the instrument via a (nano-)electrospray emitter. Gaseous ions are produced at the source and are resolved based on *m/z* in the mass analyser. Tandem MS experiments can be performed in a single mass analyser or more than one analyser may be used (either of the same or a different type). Ions are detected and this information recorded by a data system. The blue box indicates parts of the mass spectrometer operated under high vacuum.

1.2.1.1 Electrospray ionisation

The first stage of MS analysis is the generation of ions, with the requirement for "soft" ionisation techniques being critical for non-destructive ionisation of large biomolecules, such as peptides. Electrospray ionisation (ESI) is one such "soft" ionisation technique, first demonstrated for the transfer of large molecules onto the gas phase by Fenn *et al.* in the 1980s, and for which he was awarded the Nobel Prize[48, 49]. Further development of the technique allowed much lower flow rates to be used, whilst providing increased desolvation and ionisation efficiency, so called nano-ESI[50, 51]. ESI is performed under atmospheric pressure, with the application of a strong electric field and high temperature to a capillary from which the liquid sample is sprayed resulting in the formation of ions in the gas-phase. ESI results in the formation of multiply charged ions in the form of protonated adducts $[M+H]^{+}$, $[M+2H]^{2+}$, $M+nH]^{n+}$. There are two models used to describe the generation of these gas-phase ions from sprayed droplets: the ion evaporation model and the charge residue model (Figure 1.5).

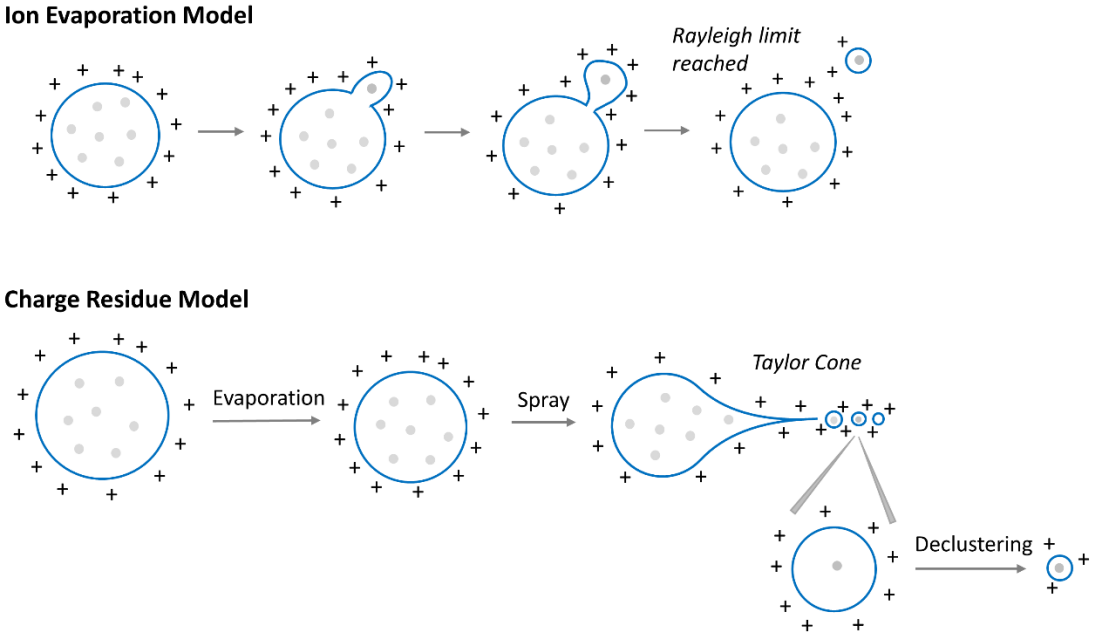


Figure 1.5 Gaseous ion formation by electrospray ionisation (ESI), described by either the ion evaporation model or charge residue model. Grey circles represent peptides, with blue outlines indicating droplets formed during ESI.

In the ion evaporation model droplets shrink by evaporation as a result of heat or a flow of gas. As a droplet shrinks, electrostatic repulsion increases up to the point at which the repulsion exceeds the surface tension of the droplet. At this point, known as the Rayleigh limit, solvated ions are expelled from the droplet[52, 53] . In the charge residue model, a highly charged droplet shrinks by evaporation until a characteristic Taylor Cone shape is formed. From the tip of the Taylor Cone, smaller highly charged droplets are emitted, with this process repeating itself until a droplet containing only one analyte ion is formed. This analyte ion is transferred to the gas phase by further solvent evaporation and declustering[50, 54]. The current consensus regarding these two models is that larger molecular ions (>1000 Da) are generated in line with the charge reside model, whilst smaller ions are more likely to be generated according to the ion evaporation model[55].

1.2.1.2 Mass analysers

Once gaseous ions enter the mass spectrometer they are guided to a mass analyser, whereby their *m/z* is determined. There are many different types of mass analyser, each performing *m/z* separation in a different way, although generally using a magnetic or electric field, or both. The two types of analyser found in the MS instrumentation used in this work are the ion trap and the Orbitrap, which are discussed in more detail below.

Ion trap

An ion trap mass analyser uses an oscillating electric field to store ions. There are two types of ion trap: the 3D (or Paul) ion trap, and the 2D (or linear) ion trap (Figure 1.6). First described in 1960[56], a 3D ion trap consists of a circular electrode with two ellipsoid electrodes at the top and bottom. Application of an RF voltage to the ring electrode whilst the end-cap electrodes are held at ground potential results in generation of a 3D quadrupolar field, which traps ions in a 3D trajectory, similar to a figure of eight. Helium gas in the trap removes excess energy from the ions meaning they oscillate at the centre of the trap where the quadrupolar field is more homogeneous, therefore giving a more uniform ejection process and improved performance[57, 58].

To perform mass analysis using an ion trap, all ions across the desired mass range are simultaneously trapped then sequentially ejected in order of mass by ramping the RF voltage applied to the ring electrode. Increasing the RF voltage causes the trajectory of ions to become unstable meaning they are expelled from the trap through the end cap

electrode to the detector. One problem with performing mass analysis in this way is that as the trajectories become unstable, some ions will be closer to the end cap than others, meaning ions with the same *m/z* may leave the trap at different times. To overcome this, mass analysis can instead by performed by resonant ejection. Here, the end-cap electrodes are supplied with an RF voltage which matches the frequency of oscillation for ions of a given *m/z*. This resonant excitation increases the amplitude of the oscillations causing the ions to be ejected. All ions of a given *m/z* will oscillate as a group so are ejected simultaneously when the resonant frequency is applied. By this principle ions of a desired *m/z* can also be selected to remain in the trap, whilst all others are ejected. If the frequencies applied to the end-cap electrodes are resonant with all ions not required, they will all be ejected leaving the ions with the desired *m/z* in the trap where they can undergo fragmentation for MS/MS analysis.

A linear ion trap consists of four quadrupoles with lenses at either end, meaning ions are confined in the radial dimension by a quadrupolar field and the axial direction by an electric field. Ions fly along the trap between the end electrodes, whilst oscillating due to the RF potential applied to the rods. As with the 3D trap, helium gas is used for cooling. Mass analysis is also performed in the same way as the 3D ion trap, either by increasing the voltage to cause ions to become unstable and be ejected from the trap, or by resonant ejection. Again, ions with a single *m/z* can be isolated in a linear ion trap to undergo fragmentation. When ions are ejected from a linear ion trap they leave via slots in two opposite rods. A detector at each of these slots means that all ions that leave the trap are analysed[59, 60]. Linear ion traps also have a higher ion storage capacity than 3D ions traps as a result of having a larger volume, and ions being focused along a line rather than a single point. Ion capacity is important as too many ions in the trap can cause space-charging effects, whereby ions nearer the edges of the trap act a shield and modify the field acting on the ions towards the centre, thus leading to decreased resolution and errors in mass measurement. In both types of ion trap this can be overcome by setting the automatic gain control (AGC) which is the maximum abundance of ions allowed in the trap[61].

Fragmentation reactions, by CID or ETD can be performed in both 3D and linear ion traps. By applying a frequency that is resonant with the isolated ions, only these ions undergo fragmentation, i.e. product ions do not fragment further as they are not in resonance. Product ions of interest can be isolated to perform additional rounds of fragmentation;

known as MS^n analysis. One limitation of performing fragmentation in an ion trap is known as 'the one-third rule'. This states that product ions with *m/z* less than one-third of the isolated precursor will have unstable trajectories and hence will be ejected before they are detected/analysed.

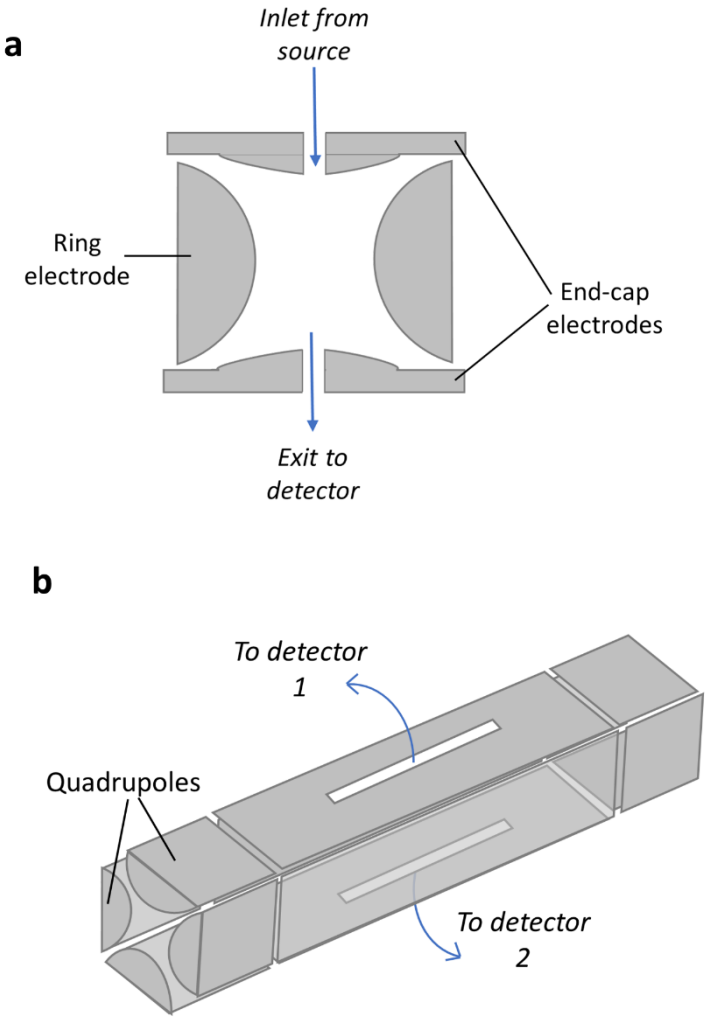


Figure 1.6 Ion trap mass analysers. a) 3D (or Paul) ion trap whereby application of an RF voltage to the ring electrode whilst the end-cap electrodes are held at ground potential results in generation of a 3D quadrupolar field, trapping ions in a 3D trajectory. **b)** 2D (or linear) ion trap whereby ions fly along the trap between the end electrodes, whilst oscillating in the radial dimension due to the RF potential applied to the four quadrupoles.

Orbitrap

The Orbitrap mass analyser was developed by Makarov in the 1990s[62, 63]. It consists of an outer 'barrel-shaped' electrode cut into two equal parts, and an inner electrode shaped like a spindle (Figure 1.7). Ions are injected as short packets into the Orbitrap, from the 'C-trap', at a position offset from its equator and perpendicular to its z-axis, which runs along the spindle electrode. Each ion packet contains ions of a single *m/z*, which immediately start coherent axial oscillations without the need for additional excitation. Stable ion trajectories in the Orbitrap involve an orbiting motion around the central electrode and simultaneous oscillations in the z-direction, along the length of the inner electrode. The frequency of oscillation along the z-axis is directly linked to the *m/z* of an ion, and is independent of the ions initial position or energy. The current induced by these oscillating ions is detected on the outer electrodes, a process termed image current detection. This time-domain signal is converted using a Fourier Transform into a mass spectrum. The main advantages of an Orbitrap mass analysers are the high resolution and high mass accuracy that can be achieved, plus a larger trapping volume compared to a 3D ion trap. The latest commercial Orbitrap instruments can deliver resolution of up to 1,000,000 FWHM at *m/z* 200, with sub-1 ppm mass accuracy.

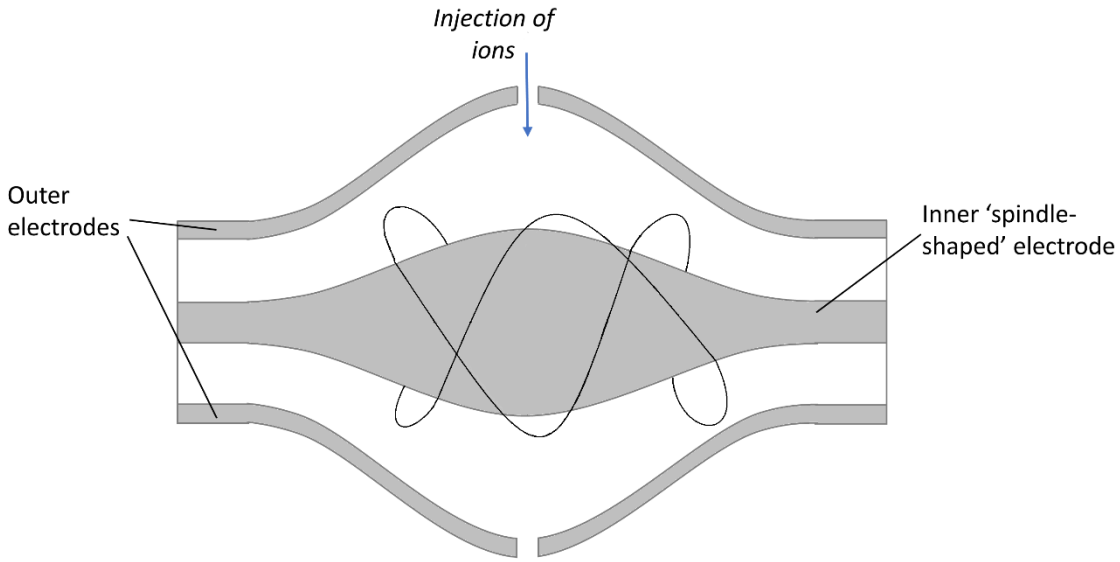


Figure 1.7 Orbitrap mass analyser. Ions oscillate around the inner electrode and axially along the length of it, with the frequency of these axial oscillations being dependent on *m/z.* The outer electrodes measure the current induced by oscillating ions, and a Fourier Transform converts the signal to a mass spectrum

1.2.1.3 Fragmentation

As previously discussed, tandem MS analysis is typically used for proteomics experiments. All ions across a designated m/z range are detected (MS1), and then precursor ions are isolated and fragmented to produce product ions which are recorded in a second stage of mass analysis (MS2). A number of different fragmentation techniques can be used to fragment the precursor ions, each with its own benefits and potential shortcomings for a particular type of experiment.

Collision-induced dissociation (CID)

CID is a generally considered as a two-step process. Firstly, ions are excited by collisions with an inert gas which results in transfer of kinetic energy into internal energy. In the second stage of the process this activated ion undergoes dissociation. As the activation time is much faster than the dissociation time, energy is distributed throughout the molecule prior to fragmentation, resulting in preferential cleavage of the weakest bonds. Peptide fragmentation by CID generally occurs via charge-directed pathways, i.e. fragmentation involves a proton at the cleavage site, which can be further described by the 'mobile proton model'[64-66]. This model states that whilst protonation of a peptide will occur at the most basic sites, activation of the peptide will allow a proton to move to other less basic sites along the peptide backbone, thereby directing cleavage of the C-N bond at these positions. Fragmentation of peptides by CID generally results in b- and y-type ions, arising from cleavage of the amide bond with the charge retained by either the amino-terminal fragment (b-ions) or carboxy-terminal fragment (y-ions). Nomenclature associated with peptide fragmentation is shown in Figure 1.8[67, 68].

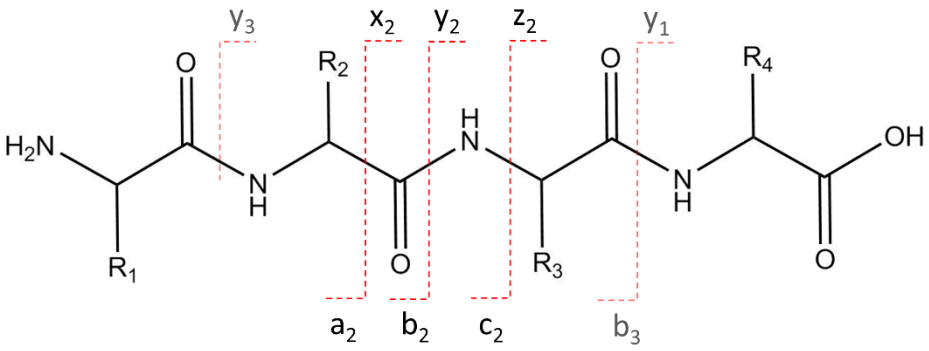


Figure 1.8 Peptide fragmentation nomenclature. Ions are labelled consecutively from the amino terminus a_n, b_n and c_n, and from the carboxy terminus x_n, y_n and z_n, where n is the number of amino acid R groups the ion contains. Cleavage of amide bonds results in b-ions when the charge is retained by the amino terminal fragment and y-ions when it is retained by the carboxy terminal fragment. Cleavage of the N-C_α bond results in c-ions when the charge is retained by the amino terminal fragment and z-ions when it is retained by the carboxy terminal fragment. A loss of C=O from a b-ion can account for the occurrence of a-ions. Double backbone cleavage, usually a result of a combination of b- and y-type cleavage produces internal fragment ions. A combination of a-type and y-type cleavage results in formation of an immonium ion.

Higher-energy collisional dissociation (HCD)

Whilst traditional CID is typically associated with lower resolution ion-trap based instruments, HCD was developed to be used alongside high resolution Orbitrap detection[69]. HCD fragmentation occurs in a collision cell containing an inert gas that is supplied with a high RF voltage. Compared to CID, the higher energy dissociations associated with HCD result in a wider range of fragmentation pathways. The generation and detection of immonium ions and other secondary fragments can aid in sequence determination and detection of post-translation modifications. HCD is considered to be similar to 'beam-type' collisional activation, producing similar spectra to those generated by CID in triple quadrupole and Q-TOF instruments. Fragmentation spectra of this type are generally dominated by y-ions, with b-ions further fragmenting to a-ions or smaller species[70].

Electron transfer dissociation (ETD)

ETD was developed in 2004 as a way to access the alternative fragmentation mechanisms of electron capture dissociation (ECD), for which an FT-ICR mass spectrometer is required[71]. Dissociation mediated by electrons (ETD and ECD) is generally thought to be a non-ergodic process, meaning that the activation occurs on a rapid timescale so bond dissociation

occurs without prior randomisation of internal energy. Therefore, fragmentation is not driven by proton localisation or bond strength, making it particularly attractive for the analysis of peptides with post-translational modifications, as the labile covalent modification remains attached to the original modified amino acid residue.

ETD is achieved by first generating reagent anions by chemical ionisation (CI). An inert gas is bombarded with 70 eV electrons to generate positively charged ions plus a population of electrons with near-thermal kinetic energy. These electrons are captured by the reagent gas, typically fluoranthene, to produce radical anions. Electron transfer occurs from the radical anion reagent to protonated peptide ions. ETD induces cleavage of the N-C_α bond, with the product ions formed termed c- and $z^{\cdot}$- type ions; the charge can be retained by either the amino-terminal fragment (c-ions) or carboxy-terminal fragment ($z^{\cdot}$-ions).

ETD is widely considered to be a complementary technique to CID[72-75]. For peptides with charge states greater than 2+, ETD can outperform both CID and HCD in terms of sequence coverage and thus peptide ion scores. A potential problem with ETD fragmentation is the occurrence of non-dissociative electron transfer, i.e. non-covalent interactions bind c- and z-ions together and thus they are not observed as ETD product ions in the mass spectrum. However, this lack of ETD product ion dissociation can be overcome by the use of supplemental activation, where low energy CID is applied following the ETD reaction, thereby breaking apart these non-dissociated ion pairs. This technique (termed ETcaD) results in a significant improvement over ETD fragmentation alone, especially for doubly charged peptides[76]. A further development of ETcsD is the integration of both ETD and HCD fragmentation to produce a mixed spectrum of b/y and c/z ions for a given precursor. The application of HCD following ETD results in fragmentation of the unreacted and charge reduced precursor ions to produce information rich fragment ion spectra[77]. Higher sequence coverage and improved phosphosite localisation are achieved with this method (termed EThcD) compared to HCD alone[78] with many benefits demonstrated in using this technique for analysis of peptide and protein post-translational modifications[79-81].

1.2.1.4 Thermo Orbitrap Fusion

The Thermo Orbitrap Fusion mass spectrometer used in this work offers an extremely flexible arrangement of mass analyser and fragmentation techniques[82, 83], with a number of

features able to be exploited to perform high quality phosphoproteomics analysis. The tribrid Orbitrap Fusion incorporates a quadrupole mass filter, linear ion trap and Orbitrap, as shown in Figure 1.9. As one of the most advanced mass spectrometers currently available for proteomics, the Orbitrap Fusions is capable of achieving 500,000 resolving power at *m/z* 200.

Operation of the Orbitrap and ion trap mass analysers can be parallelized to maximise the time available for MS1 and MS2 analysis. Isolation can be performed with the quadrupole mass filter or the ion trap mass analyser. Fragmentation can be performed using HCD, CID, ETD, ETcaD and EThcD, with these available for any level of MS^n analysis. The precursor and fragment ions can subsequently be detected using either the Orbitrap or ion trap mass analysers, representing the first time detection of HCD fragment ions by ion trap has been possible in a commercial instrument. Additionally, rather than a traditional Top*N* approach typically used for bottom-up proteomics, the 'TopSpeed' mode implements a defined cycle time during which the maximum number of MS/MS scans are performed. There are a vast number of combinatorial possibilities for tandem MS available on the Orbitrap Fusion platform. This large degree of flexibility allows for many variations in experimental design to suit a particular type of experiment.

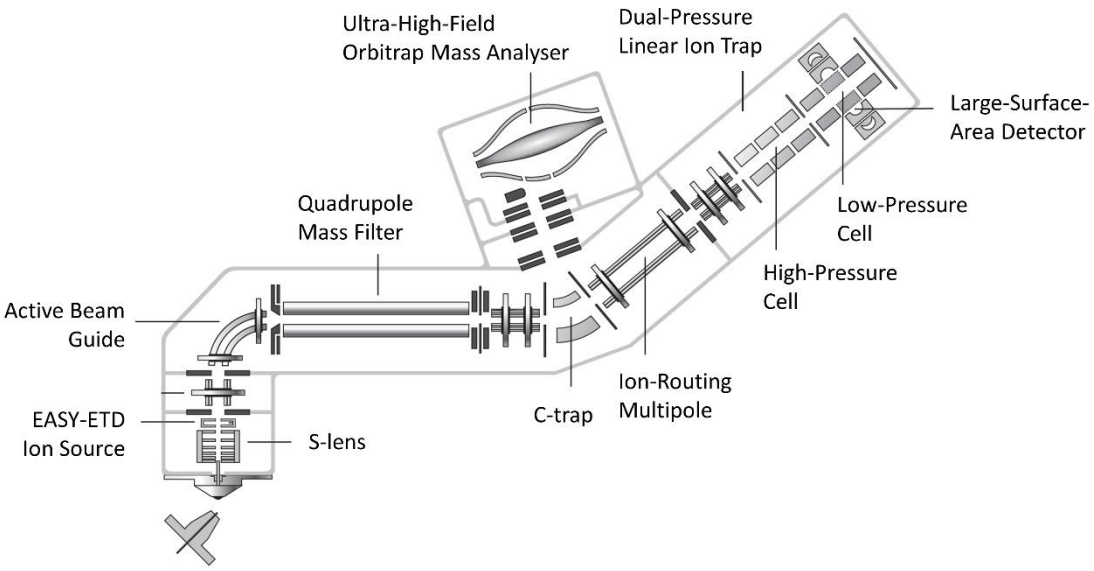


Figure 1.9 Thermo Orbitrap Fusion mass spectrometer. Includes a quadrupole mass filter, linear ion trap and Orbitrap for sensitive, rapid and flexible analysis

1.2.2 Mass spectrometry-based proteomics

MS-based proteomics is used to reveal a wealth of information about a huge variety of biological systems, from yeast to bacteria, from plants to humans[84-87]. MS-based proteomics is able to answer questions about protein structure[88], protein-protein interactions[89, 90], spatial distribution of proteins within different tissues and localization to subcellular organelles[91, 92], and post-translational modifications[93-95]. Additionally, in-depth analysis of whole proteomes provides valuable information for systems biology.

In a 'bottom-up' approach to proteomics, as illustrated in Figure 1.10, proteins are extracted from the cells, tissue or bodily fluid of interest. These proteins are then reacted with a reducing agent (e.g. dithiothreitol, DTT) to break disulphide bridges, and an alkylating agent (e.g. iodoacetamide) to prevent cysteine residues from reforming disulfide bonds, before being digested using a protease to generate a mixture of peptides. The most commonly used protease is trypsin which selectively cleaves at the C-terminus of arginine and lysine residues (except if followed by proline). At this point an enrichment step can be conducted if required, to isolate peptides of interest (e.g. phosphopeptides), or fractionation strategies, such as ion exchange or high pH reversed phase chromatography, may be carried out to increase depth of proteome coverage. The resulting peptide mixture is then subjected to liquid chromatography (LC), where peptides are separated on the basis of their hydrophobicity. Typically, peptides bound to a C18 column are eluted by increasing the concentration of acetonitrile. The LC system is coupled in-line with a mass spectrometer via ESI, and the peptide ions are analysed by tandem MS as they elute from the LC column. Tandem mass spectra are searched against a database of all possible protein sequences in the sample with specially designed algorithms, to infer peptide identities and consequently the presence of an expressed gene product (protein identity). Further details regarding analysis tandem MS data can be found in *section1.4.1*. In the alternative 'top-down approach', intact proteins are analysed rather than their constituent peptides. Unlike 'bottom-up', 'top-down' proteomics reveals information about different protein isoforms, including phosphoforms, which is inevitably lost using the 'bottom-up' approach[96-99].

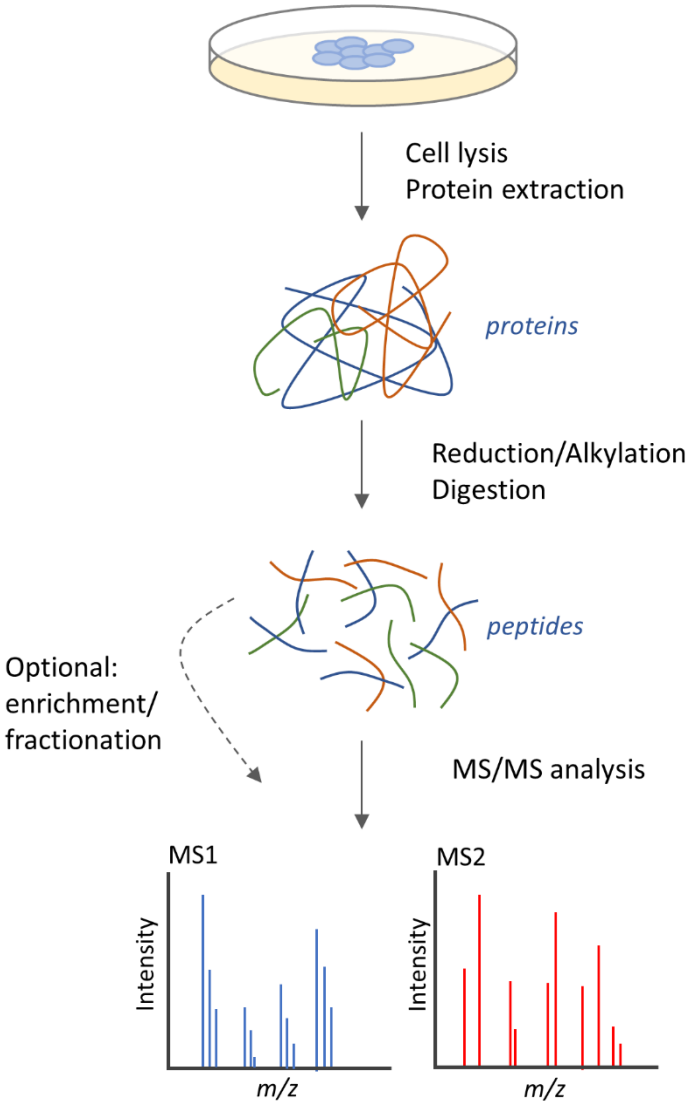


Figure 1.10 Bottom up proteomics workflow. Following extraction from the sample of interest, proteins are digested into peptides before undergoing LC-MS/MS analysis.

Bottom-up proteomics experiments can be further divided into quantitative or qualitative analysis. Qualitative experiments are designed to reveal the identity of the proteins in a biological system, which may include any PTMs the proteins have, whilst in a quantitative proteomics experiment the purpose is to determine the levels of protein expression and how they change over time or space, in various cellular states or in response to external stimuli.

Absolute quantification reports an absolute value for a given protein, for example in protein copies per cell, whilst relative quantification defines the degree of change for an individual protein between two or more conditions. Furthermore, quantitative experiments can be performed using either isotope labelling (labelled) or label-free approaches. Label free approaches arguably offer greater flexibility in experimental design, with comparison possible across many samples simultaneous without the requirement for extra experimental steps, although the can suffer from greater experimental error as samples must be processed independently. Label free quantification can be achieved by spectral counting, which is based on the observation that more abundant peptides will produce a higher number of MS/MS spectra[100] or by using peak area as a measure of peptide abundance[101]. An alternative approach termed 'Hi3' allows for absolute label-free quantification using the fact that the average abundance of the three most intense tryptic peptides is proportional to protein amount[102].

Labelling for quantitative proteomics experiments can be metabolic, for example using SILAC (stable isotope labelling by amino acids in cell culture)[103], or by chemical methods such as iTRAQ (isotope tags for relative and absolute quantification) or TMT (tandem mass tags)[104, 105]. Briefly, for a SILAC experiment cells are grown in a medium containing either 'heavy' or 'light' amino acids. For the 'heavy' culture the $^{13}C_6$/$^{15}N_7$ stable isotope-labelled versions of arginine and/or lysine are often used, whilst the 'light' culture will contain naturally occurring isotopes. The samples can be mixed immediately after harvesting of the cells which means any variability in the workflow applies equally to both samples. The 'heavy' and 'light' peptides will co-elute from the LC, but will demonstrate a mass shift in the mass spectrometer as a result of the labelled arginine/lysine residue. The ratios between the 'heavy' and 'light' peptides reflect the relative changes in abundance between the samples. Chemical incorporation of isotope tags, either by iTRAQ or TMT, is typically performed at the peptide level. Peptides from different samples are labelled with the

isobaric stable isotope tags, which consist of a reporter, balance and reactive region, and the samples are then mixed. In the MS1 spectra peptides appear a single precursor, but upon fragmentation the reporter ions are released from the tag, each of which has a distinct *m/z* value due to different combinations of ^{13}C and ^{15}N isotopes. The abundance of each reporter ion provides quantitative information regarding the peptide (and subsequently protein) abundance between samples. The TMT technology, alongside high-resolution mass spectrometry is able to provide up to 11-plex multiplexing of samples.

Using data dependent acquisition (DDA) methods, whereby precursor ions are sequentially selected for tandem MS can prove problematic for quantitative proteomics. These methods are somewhat stochastic, meaning the same peptide may not always be selected for MS/MS across multiple runs and thus 'missing values' can impede quantification of all peptides/proteins of interest in a sample. As an alternative approach data independent acquisition (DIA) methods are available. Here, all precursor ions at a given moment in time are simultaneously fragmented and the resulting mass spectrum containing all fragment ions from all precursor species is recorded. Such a method can be implemented on Waters TOF instruments, and is termed MS^E [106]. Alternatively, rather than fragmenting the entire mass range smaller mass windows (e.g. 25 Da) can be isolated and fragmented in a sequential manner until the whole mass range in covered. This technique is called SWATH, and is implemented on Sciex TripleTOF and Q-TOF instruments[107].

Defining protein PTMs, whether in a qualitative or quantitative approach, is an important area of MS based proteomics research. There are more than 300 different types of PTM, including phosphorylation, glycosylation, acetylation, ubiquitination and many more[108]. Any given gene product can be present in several modified forms, with different combinations of sub-stoichiometric modifications giving rise to a high degree of heterogeneity within the proteome. This makes comprehensive characterisation of post-translationally modified proteins particularly challenging.

1.3 Phosphoproteomics

Phosphoproteomics is a branch of proteomics specifically concerned with examining the phosphorylation state of proteins. Early phosphoproteomics was low throughput, usually only able to study the phosphorylation of one protein at a time. Some of the first MS-based phosphoproteomics experiments used a technique known as differential peptide mass mapping. Enzymatic dephosphorylation of peptides was monitored by MS, with a mass shift of 80 Da between peptides signals before and after treatment with alkaline phosphatase indicative of a peptide having previously contained a phosphorylated residue[109, 110]. To study even tens of phosphoproteins from a single experiment typically required 2D gel separation of proteins followed by individual mapping of phosphopeptides from each spot[111].

Identification of phosphopeptides from fairly complex mixtures could be performed with tandem quadrupole mass spectrometers. Using a "neutral loss scanning" method a mass offset of 98 Da between two quadrupole mass analysers means that only ions which lose H_3PO_4 reach the detector, and hence these phosphopeptides are selected for further sequence determination[112]. The "precursor ion scanning" method works in negative ion mode; ions with *m/z* 79 formed by release of PO_3^- from phosphopeptides are detected, with the corresponding precursor ion subsequently selected for MS/MS in positive ion mode[113, 114]. The low abundance of phosphopeptides in complex mixtures meant large quantities of sample were required for such analyses, and even then identification of the modified site remained a challenge.

Early improvements in LC-MS/MS based proteomics strategies also increased throughput of phosphoproteomics experiments, but even in the early 2000s only a few hundred modified peptides could be identified from a single sample, despite extensive pre-fractionation or phosphopeptide enrichment[115, 116]. In recent years, further development of protocols to enrich phosphopeptides from complex mixtures and advances in MS instrumentation and methods to enable faster, higher resolution analysis and more efficient identification of phosphopeptides, have driven the field towards high-throughput analysis. Current phosphoproteomics experiments report identification of up to tens of thousands of phosphosites from whole cell types, tissues or organisms[92, 94]. Crucially, phosphoproteomics is able to answer important questions about complex biological systems. Temporal profiles of phosphorylation have been studied in a number of systems, revealing the dynamic

nature of phosphorylation in response to external stimuli[117, 118]. Additionally, phosphoproteomics can be used to reveal e.g. aspects of tumour biology governed by phosphorylation mediated signalling, whilst both the direct and off-target effects of drugs to treat diseases such as cancer through inhibition of kinase activity can also be studied using phosphoproteomics approaches[119]. Comprehensive analysis of the phosphoproteome is thus vital for understanding human health and disease.

1.3.1 Phosphopeptide enrichment

Protein phosphorylation often occurs at sub-stoichiometric levels which means that detection of phosphorylated peptides amongst an excess of non-phosphorylated peptides is particularly challenging. Compounded by the fact that ionisation of phosphorylated peptides is reported to be suppressed by the presence of non-phosphorylated peptides, a phosphopeptide enrichment step is an essential stage of any phosphoproteomics experiment. An additional consideration for large-scale phosphoproteomics is the possible requirement to further reduce sample complexity and thereby increase depth of phosphoproteome coverage through the use of a fractionation strategy.

There are a number of strategies available for selective analysis of phosphopeptides from a complex mixture. The general principle of such strategies is to bind phosphopeptides to a solid support, allowing non-phosphorylated peptides to be discarded. A sample enriched for phosphopeptides is thus obtained, making downstream analysis of these peptides more effective.

Immobilized metal ion affinity chromatography (IMAC)

IMAC is one of the most extensively used phosphopeptide enrichment techniques. This method is based on the affinity of the negatively charged phosphate group for metal ions, such as Fe^{3+} or Ga^{3+}, which are chelated to a solid support[120, 121]. Phosphopeptide enrichment by IMAC can suffer from interference from acidic non-phosphorylated peptides, due to binding of the carboxyl groups to the metal ions. The very low pH required to fully protonate these residues can result in leaching of the metal from the IMAC material[122]. One proposed alternative is to perform an esterification reaction to convert carboxyl groups to methyl esters, but sample loss due to extensive lyophilisation and increased sample complexity as a result of incomplete reaction and unwanted side reactions can be problematic[116, 123]. Another issue with IMAC is the low tolerance for many

buffers that are typically used for preparation of biological samples, with the added limitation that IMAC in the presence of biological buffers shows a stronger selectivity for multiply phosphorylated peptides[124].

IMAC is often combined with a pre-fractionation step to reduce sample complexity prior to phosphopeptide enrichment. The use of hydrophilic interaction chromatography (HILIC) for fractionation prior to IMAC enrichment demonstrated 99% selectivity for phosphopeptides, resulting in identification of over 1000 unique phosphorylation sites from HeLa cell lysate[125]. Additionally, a combined strong cation exchange (SCX) and IMAC strategy has been used to identify ~1000 phosphopeptides in postsynaptic density samples, a significant improvement compared to using either of the techniques alone[126], whilst a similar two-step procedure consisting of SCX and IMAC resulted in identification of over 5000 phosphorylation sites from mouse liver[127].

Metal oxide affinity chromatography (MOAC)

MOAC uses immobilised metal oxide ions to selectively enrich phosphopeptides. A wide variety of metal oxides have been demonstrated to be suitable for this purpose including ZrO_2, Al_2O_3, Nb_2O_5, SnO_2, and CeO_2[128-132], but TiO_2 is by the far the most widely used. The use of TiO_2 for phosphopeptide enrichment was first described in 2004[133-135], with further developments to improve selectivity described in subsequent years[136, 137].

Selective binding of phosphopeptides to metal oxides is achieved by the affinity of the oxygen in the phosphate group for metal atoms. Performing the binding stages of the process at acidic pH minimises unwanted binding of peptides containing acidic residues; at low pH acidic residues will remain protonated. To further decrease binding of acidic no-phosphorylated peptides and thus increase selectivity, additives can be used during the binding step. A phosphate group and carboxylic acid group do not interact with TiO_2 via the same type of binding interaction (Fig 1.11). Adsorption of the phosphate group to the TiO_2 surface is via a bridging bidentate interaction, whilst carboxylic acids bind via a chelating bidentate interaction. These non-equivalent binding sites mean that additives can be chosen that will target sites that are preferred by carboxyl groups but will provide no competition for the interaction with phosphate groups. Examples of suitable additives to achieve this include 2,5-dihydroxybenzoic acid, glutamic acid, lactic acid and glycerol[136, 138-140].

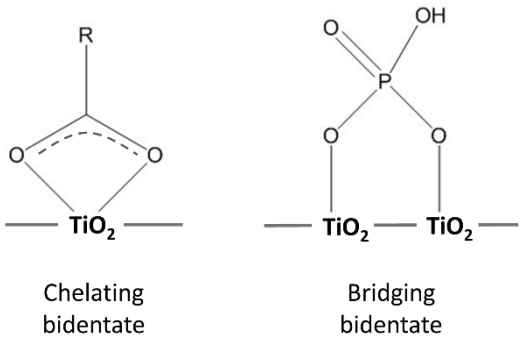


Figure 1.11 Adsorption of carboxylic acid group and phosphate onto TiO_2 by different modes

Other attempts to improve TiO_2 enrichment focus on the material used for the solid support. For example, nanoparticles have been shown to have a higher enrichment capacity than micro-particles, due to the larger surface area available for binding[141], whilst mesoporous aerogels are reported to increase the surface area further[142]. TiO_2 has been coated onto monolithic capillary columns[143], and also used to functionalize magnetic particles, which has the added advantage of increasing the ease with which the enrichment protocol can be carried out[144]. As an alternative approach, methods to enable on-target TiO_2 enrichment prior to matrix-assisted laser desorption ionisation (MALDI)-MS analysis have also been reported[145, 146].

TiO_2 for phosphopeptide enrichment has also been used alongside additional enrichment or fractionation steps to enable large scale phosphoproteomics analysis. For example, a combination of IMAC, HILIC and TiO_2 enrichment enables identification of ~4700 phosphopeptides from epidermal growth factor (EGF) stimulated HeLa cells, with quantitative analysis revealing >600 EGF regulated phosphosites[147]. An alternative approach which proposes TiO_2 as the first stage of enrichment prior to strong anion exchange (SAX) and SCX fractionation allowed for identification of ~5000 phosphosites in HeLa cells, including multiply phosphorylated peptides[148]. Additionally, an electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) strategy for fractionation with subsequent IMAC and TiO_2 enrichment provides comprehensive analysis of phosphorylation in platelet derived growth factor (PDGF) stimulated NIH 3T3 cells, allowing for characterisation of kinase substrate phosphorylation motifs[149].

Hydroxyapatite

Hydroxyapatite is a crystalline form of calcium phosphate, $Ca_{10}(PO_4)_6(OH)_2$, which was originally utilised for the separation of proteins and nucleic acids[150, 151], and is still routinely used for purification of monoclonal antibodies[152]. Phosphorylated proteins exhibit stronger binding to hydroxyapatite than non-phosphorylated proteins[153, 154], and this trend has been shown to be applicable to peptides, driving the development of hydroxyapatite as a solid support for phosphopeptide enrichment[155, 156].

Hydroxyapatite interacts with proteins and peptides through both positively charged Ca^{2+} ions which complex with carboxyl groups, and negatively charged phosphates (PO_4^{2-}) which interact with positively charged residues. The Ca^{2+} ions additionally bind to phosphoryl groups, and indeed do so with a stronger affinity than is observed for carboxyl moieties. This explains the observation that phosphorylated proteins and peptides exhibit tighter binding to hydroxyapatite than their non-phosphorylated counterparts. Hydroxyapatite enrichment routinely enables recovery of multiply phosphorylated peptides, with one proposed method enabling separation of singly, doubly and multiply phosphorylated peptides by elution from hydroxyapatite at increasing phosphate concentrations[155]. Recovery of multiply phosphorylated peptides using hydroxyapatite is particularly advantageous when compared to titanium dioxide based methods which generally favour enrichment of singly phosphorylated peptides.

The use of hydroxyapatite for phosphopeptide enrichment from relatively simple peptide mixtures has been demonstrated in a variety of formats including direct analysis of phosphopeptides by MALDI-MS without elution from the hydroxyapatite microgranules[156], incorporation of hydroxyapatite nanoparticles into monolithic pipette tips[157] and the use of magnetic hydroxyapatite clusters for a quicker and simpler enrichment protocol[158]. Additionally, hydroxyapatite has been used for phosphopeptide enrichment from complex samples in combination with MudPIT (multidimensional protein identification technology). Identification and quantification of approximately 1000 phosphopeptides was achieved from whole cell lysate, whilst the method also enabled identification of up to 4000 phosphopeptides from the amygdala region of mouse brain[159].

Calcium Phosphate Precipitation

Phosphopeptides have been found to co-precipitate with calcium phosphate at neutral pH, allowing for their separation from non-phosphorylated peptides which do not precipitate under these conditions. This technique has been applied to the analysis of complex phosphoproteomes, for example from rice[160] and brain tissue[161], either alone or in combination with an additional enrichment a step such as IMAC. More recently, trivalent lanthanide salts have been used to perform selective co-precipitation of phosphopeptides, with lanthanide cations forming strong ionic bonds with phosphate anions[162]. These precipitation methods are proposed to offer increased selectively over other affinity based methods primarily due to the fact that the precipitate can be stringently washed to remove non-phosphorylated peptides.

Immunoprecipitation

Phosphorylation of Tyr occurs at a much lower abundance than Ser and Thr phosphorylation and hence with typical phosphopeptide enrichment strategies pSer and pThr largely dominate the identified phosphosites. For this reason, pTyr is often more likely to be enriched by immunoprecipitation, with pTyr antibodies offering a more targeted enrichment strategy. Immunoprecipitation at the protein level initially suffered from an inability to identify sites of Tyr phosphorylation within pTyr proteins[163]. Additional purification of phosphopeptides by the IMAC method following precipitation of pTyr proteins improved analysis but still less than 100 pTyr sites could be identified in a single experiment[164, 165]. In 2005, a method for enrichment at the peptide level with pTyr antibodies was first described, enabling identification of over 600 pTyr sites in a cancer cell line, around 70% of which were novel[166]. Application of this method for enrichment of pTyr from a panel of cancer cell lines and tumours enabled identification of 4500 pTyr sites on over 2700 proteins[167]. Antibodies against pSer/pThr are not generally used for immunoprecipitation based enrichment on a phosphoproteome wide scale; they are targeted against a phosphoamino acid in the context of the surrounding residues and thus don't bind to all pSer/pThr sites with the same efficiency. In order to detect all pSer/pThr phosphosites by immunoprecipitation multiple antibodies would be required[168].

1.3.2 Fractionation

A variety of fractionation approaches have been employed to increase the depth of coverage in proteomics experiments. A greater overall degree of separation of peptides is achieved if two orthogonal chromatographic steps are utilised prior to MS/MS analysis, compared to a single low pH reversed phase chromatographic separation, therefore vastly increasing the number of peptide (and protein) identifications obtained for analysis of a single sample[84, 169, 170]. In phosphoproteomics experiments the combination of a pre-fractionation approach and subsequent phosphopeptide enrichment has been reported to improve the degree of enrichment and thus more comprehensive analysis of the phosphoproteome is achieved[126, 147, 159]. A number of chromatographic approaches have been exploited for both proteomics and phosphoproteomics experiments.

High pH reversed phase

Reversed phase (RP) chromatography separates peptides based on hydrophobicity; the most hydrophilic peptides elute first with peptides of increasing hydrophobicity eluting from a C18 column as the percentage of organic solvent in the mobile phase is increased. Whilst low pH RP chromatography is used as a means of peptide separation prior to MS/MS analysis, high pH RP offers an orthogonal approach that can be exploited for sample fractionation. At high pH carboxylic groups and ammonium groups will be deprotonated, changing the charge distribution and hydrophobic properties of peptides when compared to low pH where such groups will be protonated[171, 172]. Orthogonality of separation between high and low pH RP can be further increased using a concatenation approach to fraction collection; rather than adjacent fractions being combined, fractions from the start, middle and end of the gradient are combined, e.g. fractions 1, 13 and 25, and 2, 14 and 26 would be mixed together[173]. Concatenation of high pH RP fractions followed by TiO_2 enrichment resulted in comprehensive coverage of the phosphoproteome of NIH 3T3 cells, including identification of a high number of pTyr and multiply phosphorylated peptides[174].

Ion exchange

Ion exchange chromatography separates peptides based on differences in charge. Strong cation exchange (SCX) utilises a negatively charged stationary phase, such that peptides with greatest net positive charge are most strongly retained by the column. SCX has been used for effective pre-fractionation in analyses of the total proteome[84, 169]. Phosphopeptides are poorly retained by SCX, due to electrostatic repulsion of negatively

charged phosphate groups, meaning the majority of phosphopeptides elute at the start of the gradient. Nevertheless, when used either alone or in combination with an additional stage of fractionation or phosphopeptide enrichment SCX has proven to be an effective strategy for global phosphoproteomics experiments[127, 175, 176].

Strong anion exchange (SAX) facilitates peptide separation via electrostatic interactions with a positively charged stationary phase. Elution from the SAX column can be elicited by either a decreasing pH gradient[177, 178], or increasing salt concentration at a constant pH[179]. If loading of the sample onto the SAX column is performed at low pH, the negative charge of a single phosphate group is often not sufficient to overcome the electrostatic repulsion conferred by the N-terminus and the side chain of the C-terminal amino acid, resulting in poor retention of singly phosphorylated peptides. However, above pH 6 a second negative charge is acquired, making SAX fractionation of phosphopeptides a feasible strategy when high pH sample loading is employed[180].

Hydrophilic interaction chromatography (HILIC)

In HILIC, peptides interact with a neutral, hydrophilic stationary phase via hydrogen bonding. A high starting concentration of organic solvent is used in the mobile phase, with elution of peptides in order of increasing hydrophilicity elicited by increasing the polarity of the mobile phase[181]. HILIC has a high degree of orthogonality compared to reversed phase LC, making it a desirable fractionation strategy prior to standard LC-MS/MS analysis of peptides[171]. Phosphopeptides are more strongly retained by HILIC compared to non-phosphorylated peptides, with the effectiveness of HILIC for phosphoproteomics demonstrated both alone and in combination with an IMAC enrichment step[125].

Electrostatic repulsion-hydrophilic interaction chromatography (ERLIC)

ERLIC is performed using anion exchange columns with mobile phases that contain at least 60% organic solvent, leading to an additional hydrophilic interaction superimposed on the electrostatic effects[182]. The combination of these two effects results in highly effective separation of phosphopeptides from unmodified peptides[183, 184]. As is true of all electrostatic based separation methods (e.g. TiO_2, IMAC, SAX), ERLIC can suffer from co-elution of acidic peptides with phosphorylated peptides, although performing the chromatography around pH 2 is fairly effective at overcoming this problem[180, 185].

Overall, there are many strategies available for enrichment and fractionation of phosphopeptides, with a combination of approaches often offering the most in-depth coverage of the phosphoproteome. Many of these techniques rely on acidic conditions to elicit phosphopeptide enrichment and/or peptide fractionation, and therefore they are not necessarily suitable for analysis of acid-labile phosphorylation, such as pHis. Further developments of these methods are therefore required to access this unexplored area of the phosphoproteome, represented by acid-labile modifications.

1.3.1 Fragmentation of phosphopeptides for identification by tandem MS

Even once phosphopeptide samples are effectively enriched and/or fractionated, the MS stage of a phosphoproteomics experiment can still provide considerable challenges. The fragment ions detected following tandem MS must not only be able to provide peptide sequence information but also allow for the site of phosphorylation within the peptide to be determined. If the phosphate group is lost during the fragmentation reaction (a process termed neutral loss) then the resulting fragment ion spectrum may be too poor to allow for identification of the peptide sequence and/or there will be insufficient fragment ion evidence to localise the site of phosphorylation to a specific residue within the peptide. The phosphate group neutral loss can be occur as loss of HPO_3, corresponding to 80 Da or loss of H_3PO_4, corresponding to a mass of 98 Da. Concomitant loss of water (H_2O, 16 Da) alongside loss of H_3PO_4 will result in an observed neutral loss of 116 Da.

Analysis of phosphopeptides using CID fragmentation is known to result in extensive neutral loss, where the predominant ion in the tandem mass spectrum is the peptide sequence without the phosphate group. For pSer and pThr peptides this neutral loss is characterized by loss of H_3PO_4, corresponding to a mass of 98 Da. The most likely mechanism for neutral loss is a charge directed intramolecular nucleophilic substitution (Figure 1.12)[186, 187]. The reaction is initiated by protonation of the phosphate group, with the hydrogen either coming from the mobile proton, or via a protonated neighbouring amino acid. Nucleophilic attack of the β-carbon on the side chain of the phosphorylated amino acid by the amide carbonyl group results in formation of a cyclic, five membered oxazoline structure and release of H_3PO_4. The extent of this neutral loss is dependent on the charge state of the peptide and the number of basic residues (Arg, Lys, His) it contains. If the charge state is less than the number of basic residues then a low proton mobility environment exists, meaning lower energy pathways, such as neutral loss, are favoured

over backbone fragmentation. Under a mobile proton environment, where the charge state is greater than the number of basic residues, relocation of the mobile proton(s) across several amide nitrogen atoms directs C-N bond cleavage and balances backbone fragmentation with elimination of H_3PO_4.

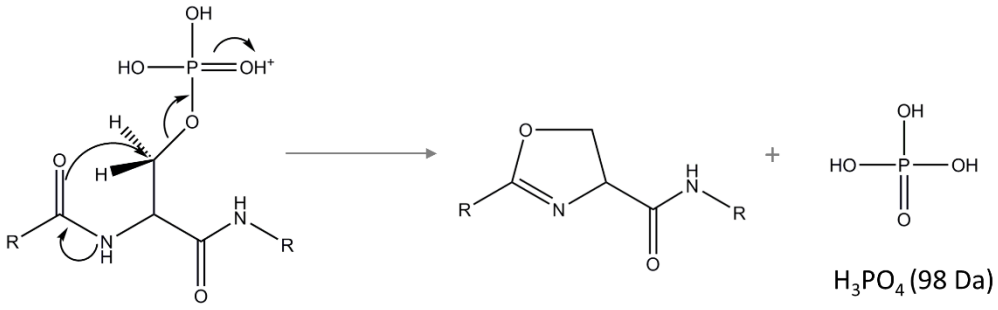


Figure 1.12 Mechanism of neutral loss from pSer/pThr peptides. Intramolecular nucleophilic substitution reaction results in formation of cyclic five-membered oxazoline structure and loss of H_3PO_4.

The aromatic ring of pTyr residues prevents loss of H_3PO_4 via the previously described mechanism. Within the C-O-P structure of a pTyr residue the C-O bond is stronger than the same bond of phosphorylated aliphatic amino acids due to stabilisation by the aromatic ring. This weakens the O-P bond, therefore favouring loss of HPO_3, corresponding to 80 Da. The exact mechanism by which the loss occurs is yet to be fully established, but it is suggested that a basic moiety may be involved. The observation of a 98 Da loss from pTyr peptides can be attributed to simultaneous or consecutive loss of HPO_3 from the pTyr residue and water (18 Da) from another amino acid. Neutral loss from pTyr peptides is generally observed to a lesser extent than from pSer/pThr peptides.

Extensive neutral loss from phosphorylated peptides, and the associated reduction in backbone cleavage, results in poor peptide sequence coverage and an inability to localise the phosphorylation to a specific amino acid[187, 188]. Attempts to improve phosphosite localization by CID have been reported, including gas-phase boron derivatisation[189] and enzymatic removal of the basic C-terminal residue, which promotes backbone cleavage over neutral loss of the phosphate group[186].

In order to improve the sequence information obtained from peptides that undergo extensive neutral loss following activation by CID, an additional round of activation can be

performed. In an MS^3 experiment, the formation of a neutral loss ion in the MS2 stage of analysis triggers a further round of activation, whereby the neutral loss ion is isolated and fragmented to produce sequence specific backbone fragments. However, this results in an increased duty cycle due to the considerable increase in analysis time per peptide. To overcome this, the MS2 and MS3 stages can be combined in a multistage activation (MSA) experiment[190]. Alongside activation of the precursor ion, additional activation at the *m/z* corresponding to the neutral loss ion is performed. This results in a spectrum which contains MS2 fragment ions plus additional fragment ions from the neutral loss species, i.e. b- and y- ions minus 98 Da. There is contradictory evidence over the usefulness of these strategies for improved phosphopeptide analysis, with the effect on phosphosite localization found to be limited[191, 192]. Although suited to some experimental approaches it is likely that the capabilities of high mass accuracy MS instrumentation negate the need for CID-based MS3 and MSA techniques for high-throughput phosphoproteomics experiments[187].

HCD has proven to be useful for phosphoproteomics experiments[193, 194], including quantitative analysis using the iTRAQ technology[195]. However, whilst detection of HCD fragment ions in the Orbitrap allows for high resolution analysis, and detection of ions in the low mass range, acquisition times are slower compared to an ion trap. This has been suggested to have possible negative implications for in-depth phosphoproteomics where rapid CID-ion trap analysis may outweigh the benefits of HCD with Orbitrap detection[196, 197].

For modified peptides ETD is particularly useful, although the slower reaction time compared to CID/HCD can impact on the overall number of MS/MS spectra acquired and therefore peptides identified in a complex mixture. Strategies combining fragmentation techniques for phosphopeptide analysis are fairly common, with one example being to perform alternating CID and ETD scans for each precursor ion[74, 118, 198], a technique employed with the Bruker AmaZon mass spectrometer used in this work.

The Thermo Orbitrap Fusion mass spectrometer used in this works offers a great degree of flexibility in terms of possible fragmentation strategies available for effective identification of phosphopeptides. Phosphoproteomics experiments can be performed with CID, HCD, ETD or a combination of fragmentation types to aid in identification[199, 200]. A particularly attractive method for this work is a neutral loss triggered workflow whereby detection of a

98 Da neutral loss (indicative of a phosphorylated peptide) in the MS2 spectrum triggers the precursor ion to undergo an additional round of fragmentation, typically using an ETD-based method. Although neutral loss triggering methods can be slow due the extra stage of MS analysis, for the purposes of this work a HCD with neutral loss triggered EThcD method is believed to provide the ideal balance between number of identifications and phosphosite confidence. Such a technique is expected to be of particular advantage for analysis of samples containing a high number of non-phosphorylated peptides, which will not necessarily benefit from EThcD fragmentation.

1.4 Data Analysis

1.4.1 MASCOT: database search engine

A vast amount of data is generated in a single LC-MS/MS experiment. From this data, both MS1 and MS2 level information are used to infer the identity of the peptides (and subsequently the proteins) in a sample. High throughput peptide/protein identification is typically performed using a search engine tool, which in the case of this work was MASCOT[201]. The basic principle of a search engine such as MASCOT is to compare the experimental masses of MS1 precursor ions and MS2 fragment ions to calculated values obtained by *in silico* digestion (and subsequent MS/MS fragmentation) of a sequence database using rules that match the experimental conditions. For example, all sequences in a database of human proteins are digested according to trypsin cleavage rules, with fragmentation of peptide ions reflecting HCD (b- and y-ions generated).

MASCOT uses a probability based scoring algorithm to enable the peptide sequence that best matches the experimental data to be identified. Essentially, the probability that the observed match between the experimental data and the sequence database occurs by chance is calculated, with a lower probability of chance being reflected by a higher score.

Current guidelines to improve reliability of protein identification data published based on database search results require the false discovery rate (FDR) to be determined[202]. Calculation of FDR is generally achieved using the target-decoy strategy[203]. According to this approach, the search is repeated with the exact same parameters except the database used is one of reversed or randomised sequences, such that there should be no matches between the experimental data set and this false (or "decoy") database. The number of matches obtained against this database is essentially an estimation of the number of false positives that are present in the results for the real (or "target") database search. FDR is thus reported as the number of false positive matches in the decoy database divided by the total number of matches in the target database.

1.4.2 Phosphosite localisation tools

The development of more sophisticated strategies for generating and analysing MS data is undoubtedly facilitating improved (phospho)peptide identification. However, for phosphopeptide identification it is equally important to have confidence in both the

sequence and the site of modification. If only one site within a peptide can be phosphorylated the site localisation is of course unambiguous, but determining with confidence the position of modification when there is more than one possibility is slightly more challenging. It is therefore beneficial to apply a localisation score to indicate confidence in a given site of modification, and indeed publication of phosphoproteomics data typically requires an assessment of such site localisation confidence.

A few search engines have an integrated method by which site localisation scores can be determined but there are also a variety of other post-search engine tools available. In general, site localisation scores can be determined in two ways. One method is to assess the chance of a given peak that allows for site determination having been matched at random, an approach which is employed by tools such as A-score[204], PTM-score (Andromeda)[117], SLoMo[205], and PhosphoRS[206]. The alternative strategy calculates the difference in search engine scores for peptide identifications with different site localisations, examples of which include MASCOT Delta score[207] and SLIP score (Protein Prospector)[208].

For the large-scale phosphopeptide analysis conducted in this project the localisation tool ptmRS was used, which is the latest version of PhosphoRS and is available as an integrated node of the Proteome Discover data analysis software[206]. The output of ptmRS is a site confidence probability for each potential site of modification within the peptide sequence. Assignment of a localisation score to each possible phosphosite is preferable to the approach taken by tools such as A-score and MASCOT Delta which report just one localisation score for an entire peptide with a given modified residue or combination of residues. A novel aspect of ptmRS is the dynamic approach to peak depth determination. Tools such as PTM-score and A-score search the 'n' most intense peaks within a bin of 100 *m/z* to identify the site determining product ions, whilst ptmRS considers the total number of extracted peaks across the entire mass range of the MS2 spectrum, whilst also taking into account the fragment ion mass tolerance. This allows ptmRS to overcome potential issues arising from uneven peak distribution in *m/z* bins. It also means ptmRS is better to suited to high-resolution data sets[200, 209]. Indeed, ptmRS was designed to be suitable for use with both high and low-resolution data, whilst both A-score and PTM-score use unit mass resolution which makes them only really suitable for low-resolution MS2 data.

Unlike FDR for peptide identifications, there is no way to readily calculate the false localisation rate (FLR) for phosphosite localisation results. Decoy sequences do not provide an accurate assessment of error as peptide identifications with incorrect site localisations are very similar to the correct matches. One possible approach is to use a synthetic library of phosphopeptides with known sites of modification, from which an FLR can be calculated and the appropriate score thresholds applied to subsequent experimental data sets. Although this strategy relies on the assumption that scores have consistent meanings when transferred from the standard data set to the experimental data set, it has nevertheless proven useful in the absence of an alternative method for calculating FLR directly from experimental data[200, 206, 209].

1.5 Challenges of phosphohistidine analysis

The study of pHis as a post-translational modification is far behind that of phosphorylated Ser, Thr and Tyr. As well as the many challenges already highlighted for phosphoproteomics experiments, the analysis of pHis presents an additional challenge. The phosphoramidate bond of pHis is acid-labile meaning the phosphate group can be lost under many of the conditions used in phosphoproteomics workflows. Hydrolysis of pHis occurs much more readily than phosphohydroxyamino acids due to the phosphoramidate bond having a -ΔG value of -12 to -14 kcal mol^{-1} compared to approximately -6.5 to -9.5 kcal mol^{-1} for a phosphoester bond[210]. Hydrolysis of both isomers of free pHis is reported to be extremely rapid under both heat and acidic conditions, typically on the time-scale of a few minutes[211, 212].

As a result of pHis instability, strategies available to study this modification in proteins are thus far fairly limited. Characterization on a protein-by-protein basis has been achieved using a membrane based method for detection of alkali-stable protein phosphorylation[213] or via a ^{31}P-NMR strategy[214]. MS-driven strategies, which have been invaluable for characterization of pSer, pThr and pTyr, would be extremely beneficial for proteome wide analysis of pHis. As already discussed, a key stage in any phosphoproteomics workflow is the enrichment of phosphopeptides from a background consisting of a large excess of non-phosphorylated peptides. Most of the currently available approaches rely on acidic conditions to elicit effective phosphopeptide enrichment, which of course is not suitable for pHis. The application of a modified IMAC approach, using Cu^{2+}, rather than the more commonly employed Fe^{3+}, has been demonstrated for enrichment of pHis peptides. However, this experiment was performed with peptides from digestion of just one protein (histidine-containing phosphocarrier protein) and is yet to be utilised for global pHis analysis[215].

For LC-MS/MS analysis, which typically employs acidic buffers for chromatographic separation of peptides, the use of shorter gradients has been investigated in an attempt to limit acid induced hydrolysis of pHis peptides[216]. Implementation of such a method will however be limited by sample complexity. Despite reports of neutral and basic buffers reducing chromatographic resolution and causing a decrease in MS signal intensity[217], Lapek *et al.*[218] describe the successful application of an LC-MS/MS based strategy for

analysis of acid-labile pHis and pAsp using ammonium bicarbonate containing buffers at pH 5. Utilising this strategy they are able to identify 20 pHis containing proteins, including the site of phosphorylation, from prostate cancer cell lines. The use of MS analysis in negative mode has also been explored, whereby reduced neutral loss in negative ion MALDI-MS analysis resulted in increased likelihood of site identification for synthetic pHis-containing peptides[219].

A further approach for pHis protein characterization is through the use of antibodies, which would permit a variety of biochemical experiments, as well as providing a strategy for pHis specific enrichment prior to LC-MS/MS analysis analogous to pTyr phosphoproteomics studies. One major problem in the development of such antibodies is that hydrolysis of pHis occurs much too quickly to elicit an immune response. To overcome this issue a number of stable analogues of pHis have been generated, for both the 1- and 3-pHis isomers[220, 221]. Several groups focused on the use of phosphoryl-triazolylalanine (pTza) as a stable pHis analogue[222, 223], with incorporation of such analogues into synthetic peptides enabling generation of sequence specific pHis polyclonal antibodies towards pHis18 in histone H4[224]. Further to this, first and second generation "pan-pHis" polyclonal antibodies against 3-pHis have been reported. Generated using either phosphoryl-triazolylethylamine (pTze) or phosphono-pyrazolylethylamine (pPye) as the hapten, these antibodies showed high affinity for a range of pHis proteins, but were limited by their cross-reactivity with pTyr[225, 226]. Most recently Fuhs *et al.*[227] incorporated the pHis analogue pTza into degenerate peptide libraries, which were subsequently used to immunize rabbits and thus the first monoclonal 1- and 3-pHis antibodies were developed. These antibodies are reported to be sequence independent with no cross-reactivity with pTyr, and their use has been demonstrated in a variety of immunological assays.

The application of monoclonal 1- and 3-pHis antibodies to an immunoprecipitation based protein enrichment, allowed over 700 pHis containing proteins to be identified by LC-MS/MS, with marked differences in proteins isolated by immunoprecipitation with each of the phosphoisomer specific antibodies. Despite this being the largest exploration of the pHis proteome to date, pHis-containing peptides were not identified, meaning that the site of phosphorylation within these proteins is still unknown. Although the development of these antibodies is a huge leap forward for the field of pHis research, there are still further advances to be made to enable confident identification and localisation of pHis sites across

the proteome. It is clear that development of new biochemical tools and novel strategies will be necessary to allow the field to determine the extent and understand the roles of His phosphorylation in mammalian systems.

1.6 Aims and objectives

The aim of this work is to develop analytical strategies for the separation and characterization of pHis-containing peptides from complex mixtures, with a view to understanding the extent of this modification in human cells. A comprehensive strategy including sample preparation, phosphopeptide enrichment, MS analysis and data processing will be established, with the aim of investigating pHis and other acid-labile phosphorylation in a human cell line. This work will endeavour to reveal the extent of the unexplored human phosphoproteome, thus gaining insight into the possible roles these modifications may play in mammalian systems.

The aims of this project will be met though the following objectives:

- Generation and characterisation of a pHis protein standard for evaluation of experimental strategies
- Assessment of existing phosphopeptide enrichment strategies for their suitability to pHis analysis
- Development of a strong anion exchange fractionation method for phosphopeptides at non-acidic pH
- Assessment of neutral-loss patterns during peptide fragmentation for improved localisation of pHis
- Application of a novel strong anion exchange strategy to HeLa cells having undergone knockdown of a known histidine phosphatase, PHPT1.
- Extensive data analysis to identify novel sites of phosphorylation on unexplored residues, with insights gained through functional annotation and motif analysis into the potential biological relevance of pHis and other non-canonical sites of phosphorylation in a human system.

Chapter 2. Materials and Methods

2.1 Sample generation

2.1.1 Preparation of histidine phosphorylated myoglobin standard

Potassium phosphoramidate was synthesised from phosphoryl chloride and ammonia according to the procedure described by Wei and Matthews[213]. In brief, phosphoryl chloride was reacted with ammonium hydroxide for 15 minutes on ice to produce ammonium hydroxide phosphate, which was subsequently reacted with potassium hydroxide at 50 °C for 10 minutes. The resulting potassium phosphoramidate salt was precipitated with ethanol and collected by vacuum filtration. Equine myoglobin (Sigma Aldrich) was phosphorylated by dissolution in 1 M aqueous potassium phosphoramidate (150 nmol/mL) overnight at room temperature. The resulting sample was buffer exchanged into 20 mM ammonium acetate using a molecular weight (10 kDa) cut-off spin filter. Briefly, sample was added to the filter, centrifuged at 10000 *x g* for 10 minutes then 500 µL 20 mM ammonium acetate added to the top of the filter before being centrifuged again. This was repeated twice more and then the concentrated sample obtained by inversion of the filter and centrifugation at 1000 *x g* for 2 minutes.

2.1.2 U2OS cell culture

U2OS cells were maintained in DMEM supplemented with 10% (v/v) fetal bovine serum, penicillin (100 U/mL), streptomycin (100 U/mL), and L-glutamine (2 mM) at 37 °C, 5% CO_2. Once 80% confluence was reached, cells were washed with phosphate buffered saline (PBS) and released with trypsin (0.05% (v/v)). Cells were centrifuged at 220 *x g* for 5 minutes, washed with PBS and lysed with 100 µL lysis buffer (8 M urea, 50 mM ammonium bicarbonate (AmBic), 1 protease inhibitor tablet (cOmplete Mini EDTA free, Roche) per 10 mL). The lysate was sonicated briefly and centrifuged at maximum speed for 20 min. Protein concentration was determined by Bradford assay.

2.1.3 HeLa cell culture and siRNA knockdown of PHPT1

HeLa cells were maintained in DMEM (Sigma-Aldrich, Dulbecco's Modified Eagle's Medium-high glucose, 4500 mg/L glucose with sodium bicarbonate, without L-glutamine and sodium pyruvate) supplemented with 10% fetal bovine serum, penicillin (100 U/mL) and streptomycin (100 U/mL) at 37 °C, 5% CO_2. To split cells, the medium was aspirated and

cells were washed with warm PBS prior to incubation with 1 mL trypsin (0.05% (v/v)) for 1 minute at 37 °C. Reaction was quenched with 1 mL supplemented DMEM. For siRNA knockdown T75 flasks at ~50% confluency were exchanged to antibiotic-free medium (DMEM supplemented with 10% fetal bovine serum). siRNA for PHPT1 (SMARTpool: ON-TARGETplus), Lamin A/C (siGENOME control) and a non-targeting pool (ON-TARGETplus Non-targeting Pool) were purchased from Dharmacon. For each flask 1 nM siRNA (1.1 µL of a 20 µM stock prepared in RNAse free water (Thermo Fisher)) and 40 µL INTERFERin (Polyplus transfection) was prepared in 4 mL Opti-mem reduced serum medium (Thermo Fisher). After incubation for 10 minutes at room temperature, the siRNA reaction mixture was added to the flasks. Cells were incubated for 48 hours at 37 °C, 5% CO_2 then trypsinized and lysed. For cell lysis, trypsinized cells were centrifuged at 220 *x g* for 5 minutes, washed with PBS and lysed with 100 µL lysis buffer (8 M urea, 50 mM AmBic, 1 protease inhibitor tablet (cOmplete Mini EDTA free, Roche) per 10 mL). Lysate was sonicated at low amplitude for 3 X 10 seconds with 1 minute in between. Protein concentration was determined by Bradford assay.

2.2 Protein analysis

2.2.1 SDS-PAGE

Cell lysates from PHPT1, Non-targeting and Lamin experiments were diluted 1:1 with 2 X sample loading buffer (0.06 M Tris-HCl (pH 6.8), 10% (v/v) glycerol, 10% (w/v) SDS, 0.005% (v/v) bromophenol blue, 0.1 M DTT), with the volume loaded for each sample resulting in equal protein loading. Samples were boiled for 5 minutes and loaded on a SDS-PAGE gel, which was run at 200 V for approximately 45 minutes. Gels were immediately used for western blotting.

2.2.2 Western blotting

Proteins were transferred to nitrocellulose membrane at 100 V for 1 hour in transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) MeOH). The membrane was blocked for 1 hour at room temperature (5% milk in TBS-T) washed with TBS-T (1 X TBS, 0.1% Tween-20; 3 X 10 minutes) and incubated with the appropriate primary antibody (PHPT1 (Santa Cruz Biotechnology, sc130229, 1:200 dilution); Lamin (Santa Cruz Biotechnology, sc6215, 1:200 dilution); GAPDH (proteintech, 1:5000 dilution)) in blocking buffer overnight at 4 °C. Membranes were then washed with TBS-T (3 X 10 minutes), incubated with the

appropriate secondary antibodies in blocking buffer for 1 hour at room temperature, washed with TBS-T (3 X 10 minutes) and developed with SuperSignal West Pico PLUS chemiluminescent substrate (Thermo Fisher) according to the supplier's protocol using X-Ray film.

2.3 Tryptic digestion and peptide desalting

2.3.1 In-solution tryptic digestion

His-phosphorylated myoglobin (200 µg) and α-/β-casein (Sigma Aldrich, Gillingham, Dorset, UK; 100 µg of each), were dissolved in 25 mM AmBic to 2 µg/µL. Cell lysates were prepared in lysis buffer, as previously described. Proteins were reduced with 3 mM DTT (in 50 mM AmBic) for 20 minutes at 30 °C, and, after cooling, free Cys residues were alkylated with 14 mM iodoacetamide (in 50 mM AmBic) for 45 minutes at room temperature in the dark. The reaction was quenched by addition of DTT to final concentration of 7 mM. The urea concentration in cell lysates was reduced to 2M by addition of 50 mM Ambic. Proteins were digested using 2% (w/w) trypsin (sequencing grade modified trypsin, Promega) at 30 °C overnight.

2.3.2 C18 StageTip desalting

StageTips were prepared with 3 discs of C18 material (Empore^{TM} Octadecyl C18, 47 mm) in a 200 µL pipette tip. Tips were conditioned by sequential addition of 100 µL MeOH, 100 µL H_2O:ACN (50:50) and 100 µL H_2O, with centrifugation for 2 minutes at 2000 *x g* to pass the liquid through the tip each time. A portion of sample (100 µL) was loaded onto the tip, centrifuged, then the flow through added to the tip and centrifuged again. The tip was washed with 100 µL H_2O and peptides were then eluted by addition of 50 µL H_2O:ACN (50:50). Eluents were dried to completion by vacuum centrifugation then resolubilised in H_2O:ACN (97:3) prior to LC-MS/MS analysis.

2.4 Phosphohistidine peptide characterisation

2.4.1 pH stability of pHis myoglobin peptides

Myoglobin peptides (1 nmol) were diluted to 100 µL in either 0.5% TFA (pH 1) or 20 mM ammonium acetate (pH 4, pH 6 or pH 9). Samples were incubated at 25 °C with shaking at 600 rpm. At timed intervals (15 minutes, 30 minutes, 1 hour and 2 hours) 5 µL of sample

was removed, neutralised (e.g. by addition of 5 µL ammonium hydroxide to samples at pH 1) and diluted to 500 fmol/µL with H_2O:ACN (97:3) for LC-MS/MS analysis with the Bruker AmaZon instrument. This experiment was performed in triplicate for each pH.

2.4.2 Heat stability of pHis myoglobin peptides

Myoglobin peptides (300 pmol) were diluted to 75 µL in 50 mM AmBic (pH 7). Samples were heated at 80 °C and 95 °C with shaking at 600 rpm. At timed intervals (30 minutes, 1 hour, 2 hours and 4 hours) 1 µL of sample was removed and diluted to 250 fmol/µL with H_2O:ACN (97:3) for LC-MS/MS analysis with the Bruker AmaZon instrument.

2.4.3 Reaction of pHis myoglobin peptides with hydroxylamine

Myoglobin peptides (600 pmol) were diluted to 20 µL with 50 mM AmBic (pH 7), and incubated with an equal volume of 500 mM hydroxylamine for 1 hour at room temperature. The sample was diluted to 500 fmol/µL with H_2O:ACN (97:3) for LC-MS/MS analysis with the Bruker AmaZon instrument.

2.4.4 Chemical modification of His-containing peptides

Tryptic peptides of myoglobin (900 pmol in 30 µL AmBic) were reacted with 30 µL of either 10 mM diethylpyrocarbonate (DEPC) in acetonitrile at room temperature or 2 mM 4-hydroxynonenal (HNE) in 10 mM sodium phosphate (pH 7.2) at 37 °C. After 1 hour an aliquot of the reaction mixture was diluted to 500 fmol/µL with H_2O:ACN (97:3) for LC-MS/MS analysis with the Bruker AmaZon instrument.

2.5 Enrichment strategies

2.5.1 Titanium dioxide enrichment

Titanium dioxide enrichment of α-/β-casein and His-phosphorylated myoglobin peptides (200 pmol) was performed using 200 µL spin tips (Protea Biosciences). For standard TiO_2 enrichment, tips were prepared by addition of 200 µL binding buffer (65% ACN, 2% TFA, saturated with glutamic acid, pH 2) and centrifuged at 2000 x *g* for 1 minute. Peptides in 200 µL binding buffer were added to the tip, centrifuged, re-loaded and centrifuged again. The resulting flow through (unbound material) was collected for analysis. Tips were washed with 200 µL each of binding buffer and two wash buffers (65% ACN, 0.5% TFA and 65% ACN, 0.1% TFA), with each fractionation collected by centrifugation. Bound peptides were

eluted by addition of 100 µL of each elution buffer (300 mM NH_3, 50% ACN and 500 mM NH_3, 60% ACN). All fractions were dried by vacuum centrifugation and reconstituted in H_2O:ACN (97:3) for LC-MS/MS analysis with the Bruker AmaZon instrument. The non-enriched peptides ("start material") were diluted to a concentration of 500 fmol/µL for LC-MS/MS analysis. Alternative binding, washing and elution buffers were also tested, as discussed further in *Chapter 3. Results I (Table 3.2).*

2.5.2 Hydroxyapatite enrichment

Hydroxyapatite (HAP) resin (5 mg; Bio-Gel HTP, Bio-Rad) was suspended in 200 µL loading buffer (20 mM Tris-HCl (pH 7.2)) and added to a Pierce spin column (Thermo Scientific). The spin column was centrifuged at 3000 *x g* for 1 minute to remove the buffer and washed with a further 50 µL loading buffer, and again centrifuged. Centrifugation of the spin column at 3000 *x g* for 1 minute was used to collect all subsequent fractions. Tryptic peptides of α-/β-casein and His-phosphorylated myoglobin (40 pmol) in 50 µL of loading buffer was added to the resin in the spin column and incubated with gentle rotation at room temperature for 30-45 minutes. The flow-through was collected and the resin washed twice with 200 µL of loading buffer, which was also collected. The resin was then washed twice with 200 µL of wash buffer (20 mM Tris-HCl (pH 7.2), 20% (v/v) ACN) which was again collected. To recover the phosphopeptides, the resin was incubated twice with 100 µL elution buffer (1.0 M K_2HPO_4 at pH 7.8, pH 7.0 or pH 6.0), rotating for 15 minutes at room temperature each time, before collecting and combining the eluent fractions. All fractions were dried by vacuum centrifugation and reconstituted in H_2O:ACN (97:3) for LC-MS/MS analysis with the Bruker AmaZon instrument.

2.5.3 Calcium phosphate precipitation

Tryptic peptides of α-/β-casein and His-phosphorylated myoglobin (50 pmol) were diluted to 50 µL in H_2O. Sodium phosphate (2 µL of 0.5 M) and ammonia water (2 µL of 2 M) were added, and the pH of the resulting solution determined to be approximately pH 10 using universal indicator paper. $CaCl_2$ (2 µL of 2 M) was added and the sample mixed using a vortex mixer for 5 minutes, followed by centrifugation at 12000 *x g* for 15 minutes. The resulting supernatant was transferred to a clean low bind sample tube. $CaCl_2$ (100 µL of 80 mM) was added to the precipitate which was briefly mixed and then centrifuged as before. The resulting supernatant was combined with that of the previous step, and the wash step repeated with a further 100 µL of 80 mM $CaCl_2$, again combining the supernatant

fractions. The precipitate was resolubilised in 10 µL of either 5% (v/v) or 0.1% (v/v) TFA and immediately transferred to a C18 StageTip for sample desalting.

2.5.4 Subtractive approaches for pHis peptide enrichment

Tryptic peptides of α-/β-casein and His-phosphorylated myoglobin (20 pmol) in 50 µL of loading buffer (20 mM Tris-HCl (pH 7.2)) were added to HAP resin (5 mg; Bio-Gel HTP, Bio-Rad) and washed as described above (*section 2.5.2)*. The resin was then washed with 100 µL of H_2O and incubated with 500 mM hydroxylamine (50 µL) for 1 hour at 37 °C or without hydroxylamine at 95 °C for 45 minutes. The eluates were collected, dried to completion and reconstituted in H_2O:ACN (97:3) for LC-MS/MS analysis with the Bruker AmaZon instrument.

Tryptic peptides of α-/β-casein and His-phosphorylated myoglobin (200 pmol) in 200 µL binding buffer (65 mM NH_4OAc, 5% MeCN, (pH 7.5)) were added to a TiO_2 spin tip (Protea Biosciences), centrifuged (2000 x *g* for 1 minute), re-loaded and centrifuged again. The tip was then washed with 100 µL of binding buffer, followed by addition of 50 µL 5% (v/v) TFA. The tip was incubated with the acid for 30 minutes then the eluate collected by centrifugation. Bound phosphopeptides were eluted with 100 µL each of 300 mM NH_3, 50% MeCN and 500 mM NH_3, 60% MeCN. All eluates were dried to completion and reconstituted in H_2O:ACN (97:3) for LC-MS/MS analysis with the Bruker AmaZon instrument.

2.5.5 Immunoprecipitation

Crosslinking of 1- and 3-pHis antibodies (2 µg) to 25 µL protein A/G magnetic beads with DSS was conducted using the Pierce Crosslink Magnetic IP kit (Thermo Scientific). U2OS cell lysate (500 µg) was diluted to 500 µL with wash/binding buffer (50 mM Tris, 30 mM sodium carbonate, pH 8) and a pre-clearing step performed by incubating the lysate with 25 µL protein A/G magnetic beads for 1 hour at room temperature. Pre-cleared lysate was then incubated with 1- and 3-pHis antibody crosslinked magnetic beads at 4 °C overnight. Beads were washed with 3 X 200 µL wash/binding buffer and once with 500 µL H_2O prior to elution with 3 X 100 µL elution buffer (100 mM trimethylamine, pH 11). Eluted proteins were dried to completion by vacuum centrifugation and re-solubilised in 25 mM AmBic for tryptic digestion.

2.6 Fractionation

2.6.1 SAX chromatography

SAX was performed using a Dionex U3000 HPLC instrument equipped with a fraction collector. Peptides from phosphorylated protein standards (25 µg each α-/β- casein and His-phosphorylated myoglobin) or digested cell lysate (2 mg) were chromatographed using a PolySAX LP column (PolyLC; 4.6 mm × 200 mm, 5 µm particle size, 300 Å) with a binary solvent system of solvent A (20 mM ammonium acetate, 10% ACN) and solvent B (300 mM triethylammonium phosphate, 10% ACN) at pH 6.0, pH 6.8 or pH 8.0. Solvent was delivered at 1 mL/min according to the following gradient: 5 minutes at 100% solvent A, then 43 minutes to 100% solvent B, and then 5 minutes at 100% solvent B before equilibration to start conditions. Fractions were collected every minute for 48 minutes, with every 3 pooled and the volume reduced by drying under vacuum to give 16 fractions in total.

2.6.2 High pH RP chromatography

High pH RP fractionation was performed using a Dionex U3000 HPLC instrument equipped with a fraction collector. Peptides from phosphorylated protein standards (25 µg each α-/β-casein and His-phosphorylated myoglobin) were separated by RP-HPLC on an Extend-C18 column (3.5 µm packing material, 3 mm x 150 mm; Agilent Technologies Ltd., Stockport, UK) using a binary solvent system of solvent A (20 mM ammonium hydroxide, pH 10) and solvent B (20 mM ammonium hydroxide in ACN:H_2O 90:10). Solvent was delivered at 500 µL/min according to the following gradient: 5 minutes at 5.5% solvent B, then 31 minutes to 34% solvent B, 6 minutes to 45% solvent B, a further 6 minutes to 100% solvent B, then 5 minutes at 100% solvent B before equilibration to start conditions. Fractions were collected every minute for 48 minutes, with every 3 pooled and the volume reduced by drying under vacuum to give 16 fractions in total.

2.6.3 SAX fractionation using StageTips

SAX buffers for step elution were prepared by mixing Buffer A (20 mM ammonium acetate, 10% ACN, pH 6.8) and Buffer B (300 mM triethylammonium phosphate, 10% ACN, pH 6.8) as follows: 'SAX1' 100:0 (A:B); 'SAX2' 95:5; 'SAX3' 90:10; 'SAX4' 80:20; 'SAX 5' 70:30; 'SAX6' 60:40; 'SAX7' 50:50 'SAX8' 30:70. StageTips were prepared with 2 discs of anion exchange material (Empore Anion Exchange, 47 mm) followed by 1 SDB disc (Empore styrenedivinylbenzene-XC, 47 mm) in a 200 µL pipette tip. Tips were conditioned by

sequential addition of 20 µL MeOH, 20 µL Buffer 2 (20 mM ammonium acetate, 80% ACN), 20 µL Buffer 1 (20 mM ammonium acetate, pH 6.8), 20 µL buffer 'SAX8' and 20 µL Buffer 1, with centrifugation at 500 *x g* for 2 minutes each time. Tryptic peptides of α-/β-casein and His-phosphorylated myoglobin (20 µg of each) were added to the tip in 50 µL of Buffer 1 and centrifuged as before, then the tip was washed with 20 µL Buffer 1 and the flow through discarded. Buffer 2 (20 µL) and 2 X 20 µL buffer 'SAX1' were sequentially passed through the tip, with the flow through combined as Fraction 1. Sequential addition of 2 X 20 µL of each buffer 'SAX2' to 'SAX 8' with centrifugation to collect the flow through was performed to obtain Fractions 2-8. All SAX fractions were made up to 100 µL with H_2O prior to desalting with C18 StageTips.

2.7 Mass spectrometry

2.7.1 Intact protein MS analysis

Analysis of intact phosphorylated myoglobin was conducted using a Waters Synapt G2-S*i* instrument. Phosphorylated myoglobin (5 µM in 20 mM ammonium acetate:ACN 50:50) was analysed using borosilicate emitters (Thermo ES 387). Spraying voltage was adjusted to 1.5 kV, sampling cone was 50 V. The time-of-flight mass analyser was set to Resolution Mode. Data was processed using Mass Lynx V4.1, with deconvolution performed using MaxEnt 1.

2.7.2 LC-MS/MS

Bruker Amazon ETD

LC-MS/MS analysis of simple peptide mixtures was performed using the AmaZon ETD ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany) arranged in-line with a nanoAcquity n-UHPLC system (Waters Ltd., Elstree, UK). Peptides were loaded from an autosampler onto a Symmetry C_{18} trapping column (5 µm packing material, 180 µm x 20 mm) (Waters Ltd., Elstree, UK) at a flow rate of 5 µL/min of solvent A (0.1% (v/v) formic acid in H_2O), trapped for 3 minutes and then resolved on a nanoACQUITY C_{18} analytical column (1.8 µm packing material, 75 µm x 150 mm) (Waters Ltd., Elstree, UK) using a gradient of 97% A, 3% B (0.1% (v/v) formic acid in ACN) to 60% A, 40% B over 60 minutes at 300 nL/min. The column effluent was introduced into the AmaZon ETD ion trap mass spectrometer via a nano-ESI source with capillary voltage of 2.5 kV. Full scan ESI-MS spectra were acquired over 150-2000 *m/z*, with the three most abundant ions being selected for

isolation and sequential activation by CID or ETD. A 1 min dynamic exclusion window was incorporated to avoid repeated isolation and fragmentation of the same precursor ion. CID was performed with helium as the target gas, with the MS/MS fragmentation amplitude set at 1.20 V, and ramped from 30 to 300% of the set value. For ETD, peptides were incubated with fluoranthene anions (ICC target 100000, max ETD reagent accumulation time 10 ms, ETD reaction time 100 ms).

Thermo Orbitrap Fusion

nLC-ESI-MS/MS analysis of cell lysate samples was performed using an Orbitrap Fusion tribrid mass spectrometer (Thermo Scientific) attached to an UltiMate 3000 nano system (Dionex). Peptides were loaded onto the trapping column (Thermo Scientific, PepMap100, C18, 300 µm X 5 mm), using partial loop injection, for seven minutes at a flow rate of 9 µL/min with 2% ACN 0.1% (v/v) TFA and then resolved on an analytical column (Easy-Spray C18, 75 µm x 500 mm, 2 µm bead diameter) using a gradient of 96.2% A (0.1% formic acid in H_2O) 3.8% B (0.1% formic acid in 80:20 ACN:H_2O) to 50% B over 90 minutes at a flow rate of 300 nL/min. A full scan mass spectrum was acquired over *m/z* 350-2000 in the Orbitrap (120K resolution at *m/z* 200) and data-dependent MS/MS analysis performed using a top speed approach (cycle time of 3 s), with HCD (collision energy 32%, max injection time 35 ms) and neutral loss triggered (Δ98) EThcD (ETD reaction time 50 ms, max ETD reagent injection time 200 ms, supplemental activation energy 25%, max injection time 50 ms) for fragmentation. All product ions were detected in the ion trap (rapid mode).

2.8 Data analysis

2.8.1 CompassXport/MASCOT

MS output files from the Bruker AmaZon instrument were converted to .mgf files using CompassXport software (Bruker Daltonic) and an in-house script. The resulting .mgf files were searched using the MASCOT search algorithm (version 2.6) against the Uniprot bovine and equine databases. Parameters were as follows: MS1 and MS2 mass tolerance of 0.6 Da; enzyme specificity set as trypsin with 2 missed cleavages allowed; carbamidomethylation of Cys set as a fixed modification; phosphorylation of Ser, Thr, Tyr and His set as variable modifications. Data was manually inspected using DataAnalysis software (Bruker Daltonic) for the presence of pHis-containing peptides, and to extract peak area values for (phospho)peptides.

2.8.2 PEAKS 7.5

Label-free quantification of myoglobin peptides in SAX fractions (either with casein or spiked into U2OS lysate) was performed following analysis with the Orbitrap Fusion tribrid mass spectrometer. Raw files were searched using the PEAKS search engine against either an in-house database created by combining the Uniprot Bovine and Equine databases (2016.02.19) or the Uniprot human reviewed database (2015.12.02). Search parameters were as follows: parent mass error tolerance of 10 ppm; fragment mass error tolerance of 0.6 Da; enzyme specificity set as trypsin with 2 missed cleavages allowed; carbamidomethylation of Cys set as a fixed modification; phosphorylation of Ser, Thr or Tyr, phosphorylation of His, Cys, Asp or Arg, and oxidation of Met set as variable modifications with the maximum number of variable modification per peptide set to 4.

2.8.3 Proteome Discoverer 1.4

Raw files acquired on the Thermo Fusion mass spectrometer (HCD-neutral loss triggered EThcD method) were converted to .mzML using ProteoWizard's msconvert tool in order to perform MS2-level deisotoping. The resulting files were then processed using Proteome Discoverer. For each raw file scans were split into those arising from HCD and EThcD events using a collision energy filter (HCD: min CE 0, max CE 34; EThcD: min CE 35, max CE 1000) to generate 2 separate .mgf files. These were searched using the MASCOT search algorithm (version 2.6) against the Uniprot Human database (2015.12.02; 20,187 sequences). Parameters were set as follows: MS1 tolerance of 10 ppm; MS2 mass tolerance of 0.6 Da; enzyme specificity set to trypsin with 2 missed cleavages allowed; carbamidomethylation of Cys set as a fixed modification; phosphorylation of Ser, Thr, Tyr and His and oxidation of Met set as variable modifications; instrument type set as ESI-Quad-TOF for HCD files and CID+ETD for EThcD files. Phosphopeptides were additionally analysed using the ptmRS node, with the 'treat all spectra as EThcD' option selected for EThcD data. A peptide FDR filter of 5% was applied and all data files were exported to .csv for further processing. Peptides with rank >1 and those classified as 'ambiguous' or 'rejected' were removed. MS2 spectra corresponding to phosphopeptides were assessed for the presence of 3 neutral loss peaks (Δ80, Δ98 and Δ116) with a mass tolerance of 0.5 Da and an intensity cut-off of 5% compared to the base peak ion and given a 'Triplet' score of 0, 1, 2 or 3 depending on the number of neutral loss peaks identified.

2.8.4 Bioinformatics

DAVID functional annotation and clustering

Function annotation and clustering was performed using DAVID (version 6.8). The list of pX proteins identified in each condition (NT siRNA and PHPT1 siRNA) was compared to a 'user-submitted' background of all proteins identified by analysis of all samples. Default settings were used. Classification stringency: medium; similarity term overlap: 3; initial/final group membership: 3; EASE: 1.0.

Enriched functional groups were classified according to P value: <0.05; <0.01 and <0.001. Groups with Benjamini adjusted P-values < 0.1 are also highlighted. The Benjamini-Hochberg method is a strategy for correcting p-values to account for multiple hypothesis testing[228]. This method first ranks all items (e.g. genes/proteins) by P-value; the adjusted P-value for a given item is then calculated by multiplying the original P-value by the total number of items divided by the rank of the given item.

Motif-X

Sequences for input into Motif-X were extracted from pX site/protein information to obtain the seven residues either side of the phosphorylated site. These '15-mers' were analysed using the Motif-X algorithm with significance set at 1E-5 (corresponds to P value of 3E-4 with Bonferroni correction). The Bonferroni correction is a method of controlling for multiple comparisons; the P-value is divided by the number of statistical tests to determine a new, smaller P-value below which results are determined to be significant.

Chapter 3. Results I: Exploring strategies for the enrichment of phosphohistidine-containing peptides

3.1 Introduction

As previously discussed in *Chapter 1. Introduction*, the aim of this work was to develop strategies suitable for the analysis and identification of acid-labile pHis-containing peptides from complex biological samples. Phosphopeptide enrichment is a key stage in the phosphoproteomics workflow, yet techniques suitable for pHis-peptide enrichment are lacking. This chapter describes the generation and validation of a pHis protein standard, and its application in the characterisation of multiple phosphopeptide enrichment strategies for their suitability for the analysis of labile pHis-containing peptides.

The phosphoproteins α- and β-casein are routinely used as standards to assess new and existing phosphopeptide enrichment strategies[136, 155, 229] as they yield a number of well characterised pSer and phosphothreonine (pThr)-containing peptides upon tryptic digestion. However, there are currently no known analogous pHis-containing proteins which would provide multiple pHis-containing peptides upon enzymatic digestion. Additionally, pHis-containing peptides cannot be synthesised by typical solid-phase synthesis strategies due to the instability of the phosphoramidate bond. However, potassium phosphoramidate can be used to chemically phosphorylate His as a free amino acid, in synthetic peptides and also within proteins[211, 217, 230, 231], and was therefore utilised to generate a pHis-containing protein which could be used as a standard for method development.

The initial strategy for this work was to assess a number of published phosphopeptide enrichment strategies for their suitability for the analysis of pHis-containing peptides. Titanium dioxide (TiO_2) enrichment is one of the most widely used techniques for phosphopeptide enrichment[136, 147, 232, 233], although the acidic conditions used to enhance selectivity of enrichment may not necessarily be suitable for pHis. Another acid-labile modification, phosphoarginine, has previously been shown to be enriched using an alternative TiO_2 based enrichment strategy at pH 4[234]. Hydroxyapatite is known to bind particularly strongly to phosphoproteins and has therefore been used as a solid support to

perform phosphopeptide enrichment[155, 156, 159]. The non-acidic conditions make this particularly attractive for enrichment of peptides containing acid-labile pHis residues. A further strategy used for enrichment of phosphopeptides is based on their precipitation with calcium phosphate under basic conditions[160, 161]. All of these methods have previously been used to enrich pSer/pThr peptides in complex samples but thus far none of them have been assessed for pHis analysis.

In addition to these potential enrichment strategies, chemical derivatisation approaches for analysis of pHis-containing peptides were also considered. The rationale for such an approach would be to chemically derivatise non-modified His residues, prior to actively inducing loss of phosphate from pHis residues. In this way, His-phosphorylated peptides could be inferred from their absence of chemical derivatisation. Diethylpyrocarbonate (DEPC) and 4-hydroxynonenal (HNE) are two possible candidates for selective His modification that were tested as part of this work[235, 236].

Finally, in response to the development of monoclonal 1- and 3-pHis antibodies[227], an immunoprecipitation experiment was conducted. In contrast to the other techniques trialled, this enrichment is conducted at the protein level. Although this technique has been reported to recover pHis proteins from cell lysates, the ability of this method to identify His phosphorylated peptides is yet to be established.

3.2 Results

3.2.1 Generation of a histidine-phosphorylated protein standard

Potassium phosphoramidate was used to chemically phosphorylate the His residues of myoglobin. The specificity of the reaction is such that Ser, Thr or Tyr residues are not phosphorylated[219] although reaction with potassium phosphoramidate can also be used to phosphorylate lysine residues[237, 238]. Myoglobin contains 11 His residues which upon phosphorylation and digestion can theoretically yield up to 6 pHis-containing tryptic peptides. Potassium phosphoramidate was synthesised in-house according to a previously reported reaction scheme (Figure 3.1). Phosphoryl chloride ($POCl_3$) was reacted with ammonium hydroxide on ice to give ammonium hydrogen phosphoramidate ($NH_4HPO_3HN_2$). This was subsequently reacted with potassium hydroxide (KOH) at 50 °C and acidified to produce potassium phosphoramidate ($KHPO_3NH_2$). Overall reaction yield was typically around 15%.

Myoglobin was reacted with potassium phosphoramidate overnight (Figure 3.2); it is reported for free pHis that shorter reaction times yield the 1-pHis isomer whilst increasing the reaction time (to at least 16 hours) results in formation of the more thermodynamically stable 3-pHis isomer[211]. Buffer exchange was performed to remove the large excess of potassium phosphoramidate and to ensure the protein was in a non-volatile solvent suitable for direct infusion into a mass spectrometer.

The deconvoluted mass spectrum obtained following intact mass analysis of myoglobin that had been reacted overnight with potassium phosphoramidate is shown in Figure 3.3. Although a small amount of unreacted myoglobin can be observed, the vast majority of signal derives from phosphorylated forms of the protein. The most abundant phosphoisomers contain three phosphate groups, with up to five phosphoforms detected. The heme-bound form of myoglobin can also be observed, again with multiple phosphoforms, albeit in much lower abundance.

Enzymatic digestion of phosphorylated myoglobin with trypsin was performed and the resulting peptides analysed by LC-MS/MS. Details of the His phosphorylated peptides detected following LC-MS/MS analysis with the Thermo Orbitrap Fusion are shown in Table 3.1.

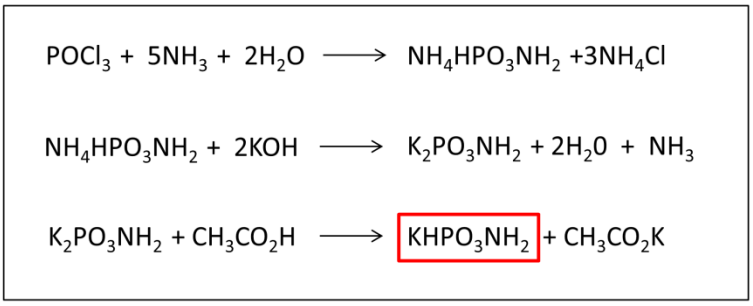


Figure 3.1 Reaction scheme for synthesis of potassium phosphoramidate ($KHPO_3NH_2$) via ammonium hydrogen phosphoramidate ($NH_4HPO_3HN_2$)

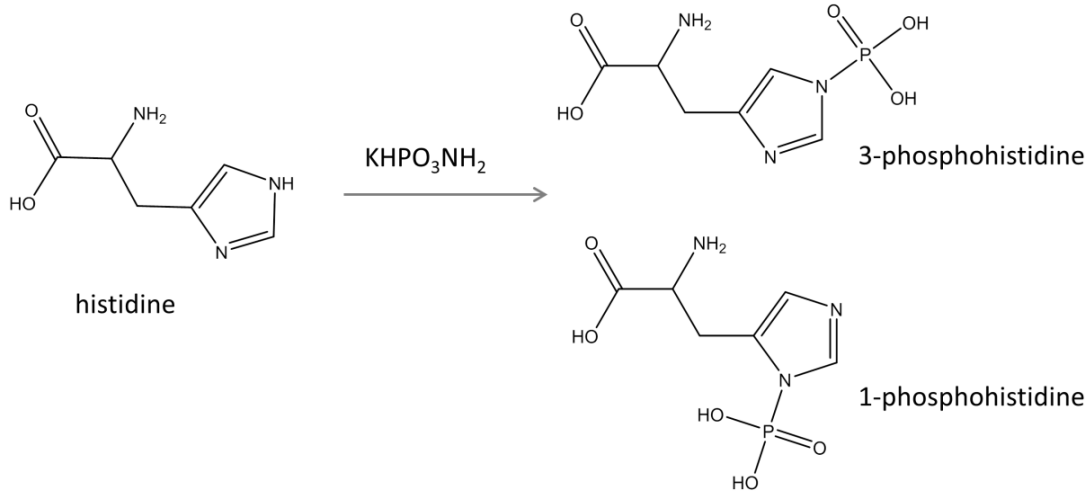


Figure 3.2 Reaction scheme for phosphorylation of histidine by potassium phosphoramidate. Reaction overnight yields the more thermodynamically stable 3-pHis isomer

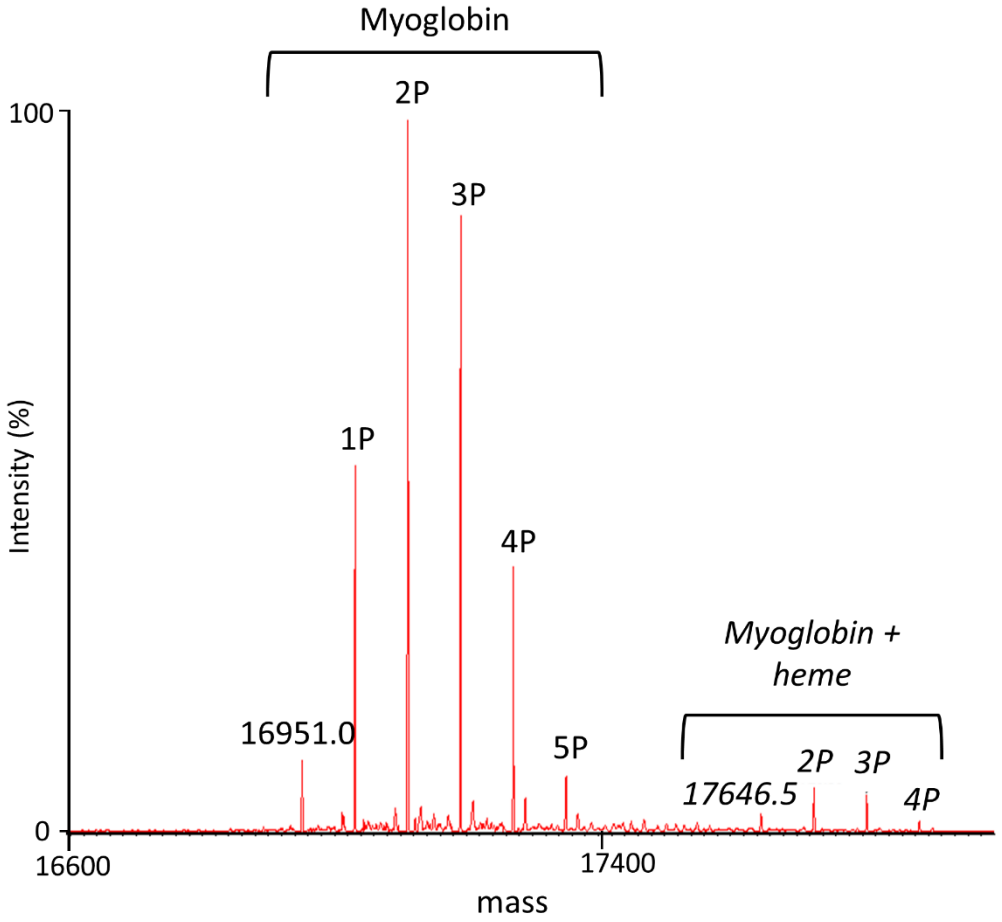
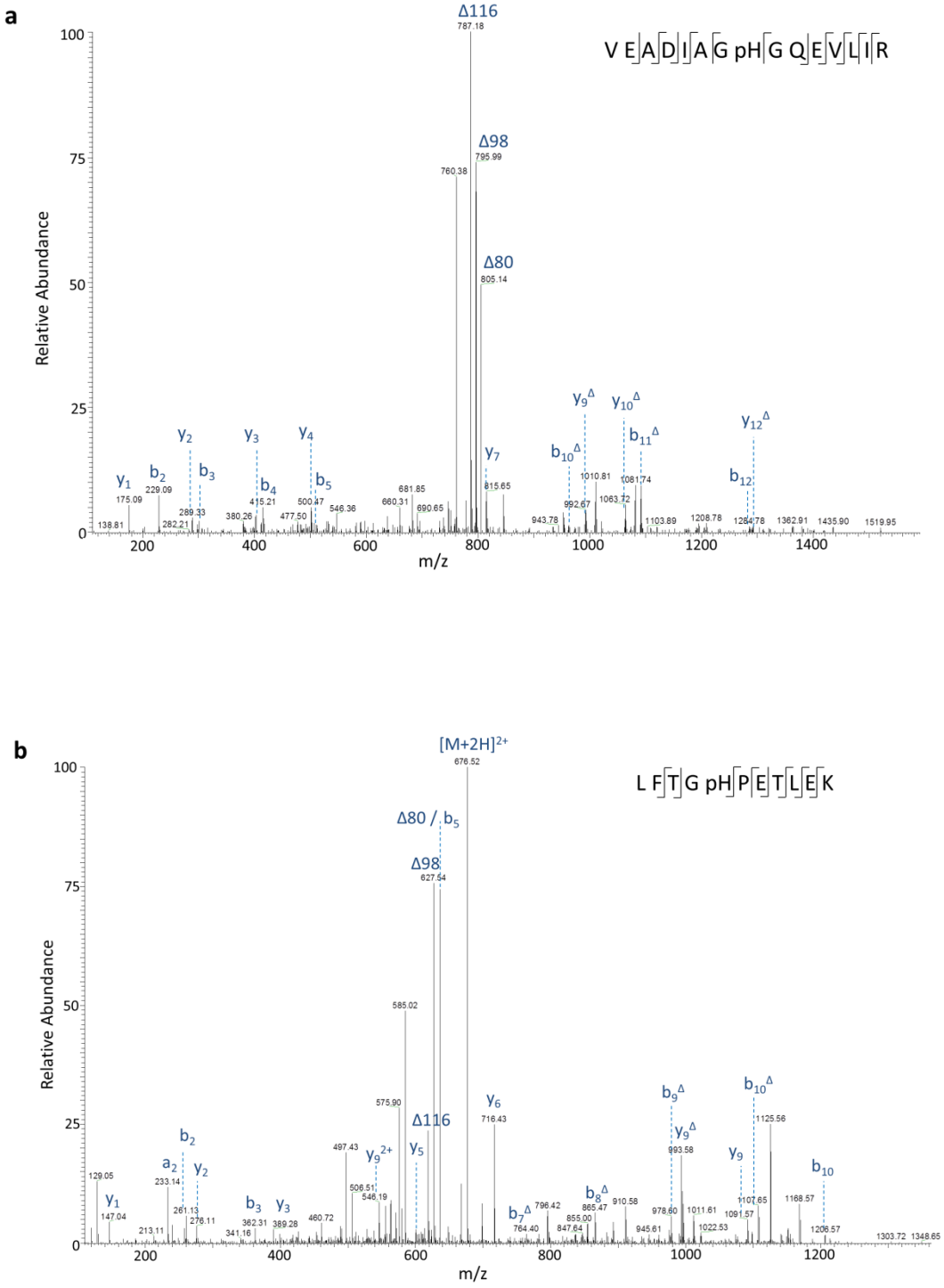


Figure 3.3 Zero charge state mass spectrum of intact phosphorylated myoglobin following overnight reaction with potassium phosphoramidate, generated by direct infusion via nanoESI into a Synapt G2-S*i* mass spectrometer. The raw mass spectrum was deconvoluted using MaxEnt. Up to five phosphate groups per intact myoglobin molecule are observed, along with the heme-bound form of myoglobin containing up to four phosphate groups.

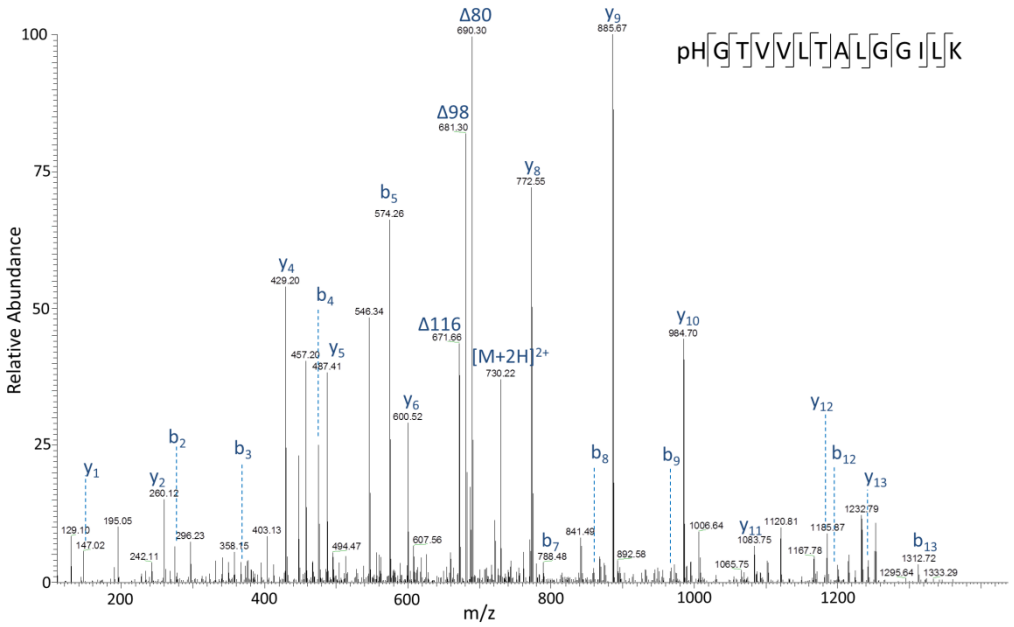
Table 3.1 Phosphohistidine-containing peptides of myoglobin generated by overnight reaction of myglobin with potassium phosphoramidate and identified following tryptic digestion and LC-MS/MS analysis with HCD or EThcD fragmentation. Phosphosite localisation was confirmed by manual annotation of the spectra.

Sequence	**Site**	**Observed *m/z***	**Charge**
VEADIAGpHGQEVLIR	His25	845.42	2^+
LFTGpHPETLEK	His37	676.32	2^+
pHGTVVLTALGGILK	His65	730.41	2^+
GpHHEAELKPLAQSHATK	His81	645.31	3^+
GHpHEAELKPLAQSHATK	His82	645.31	3^+
GHHEAELKPLAQSpHATK	His94	645.31	3^+
YLEFISDAIIpHVLHSK	His114	983.50	2^+

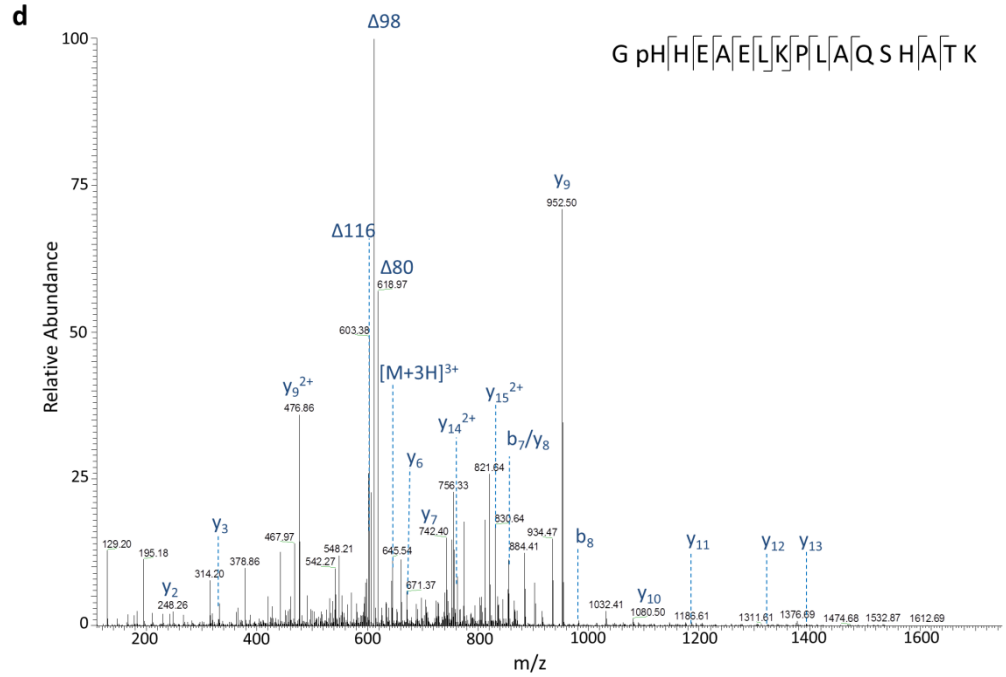
Tandem mass spectra were manually annotated in order to confirm the identity of each of the pHis-containing peptides (Figure 3.4.a-g). The five pHis-containing peptide sequences identified correspond to a total of seven pHis sites, i.e. one peptide (GHHEAELKPLAQSHATK) contains three possible His residues that could be phosphorylated. Figures 3.4.d, 3.4.e and 3.4.f indicate that there is evidence for any of the 3 possible sites of this peptide to be phosphorylated. There is no selectivity of the potassium phosphoramidate reaction in terms of which His residues can be phosphorylated, so any (or all) of these phosphoisomers could be present in the sample. Knowing which His residue is phosphorylated is not relevant for the assessment of various enrichment strategies. No evidence has been found for the presence of this peptide in a multiply phosphorylated form. It is worth noting that the non-phosphorylated equivalents of these peptides are also observed by LC-MS/MS following tryptic digestion, due presumably to a combination of the fact that some myoglobin remained unphosphorylated after reaction with potassium phosphoramidate (as previously shown by intact mass analysis, Figure 3.3) and the potential for His residues to become dephosphorylated during LC-MS/MS analysis.


a
V E A D I A G pH G Q E V L I R
Δ116
787.18
Δ98
795.99
760.38
Δ80
805.14
Relative Abundance
m/z
b
[M+2H]2+
676.52
L F T G pH P E T L E K
Δ80 / b5
Δ98
627.54
Δ116
716.43
585.02
575.90
Relative Abundance
m/z

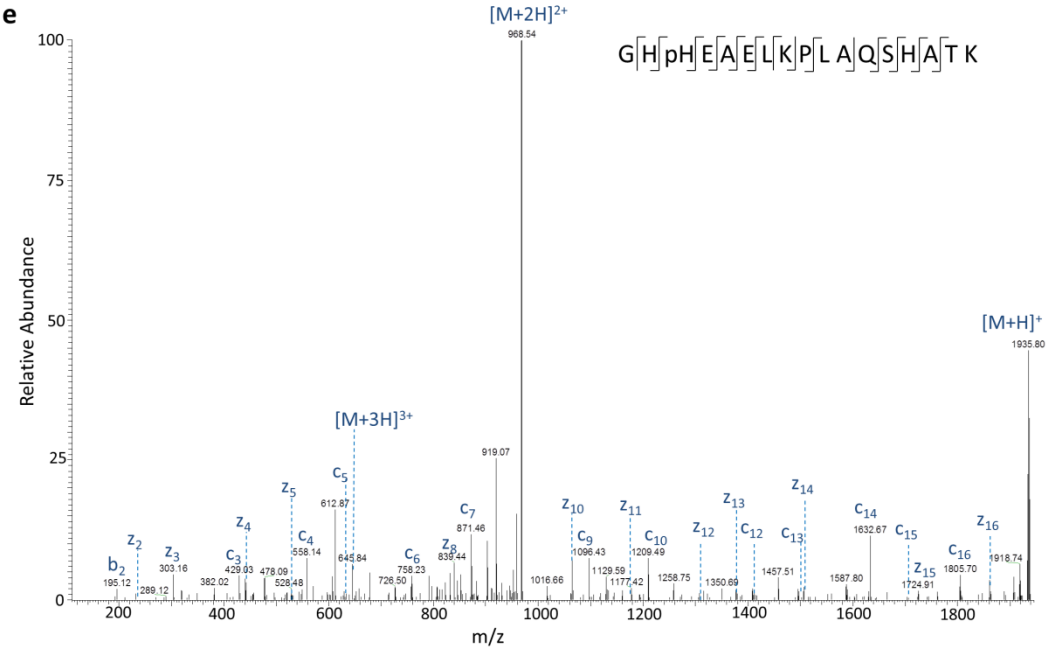
c


pH G T V V L T A L G G I L K
Δ80
690.30
y9
885.67
Δ98
681.30
y8
772.55
b5
574.26
y4
429.20
b4
546.34
Δ116
671.66
[M+2H]2+
730.22
y10
984.70
y5
487.41
457.26
y6
600.52
y12
b2
b3
b8
b9
b12
y1
y2
260.12
y13
1232.79
b7
788.48
841.49
b13
1312.72
y11
1083.75
1120.81
1185.87
1006.64
1167.78
1065.75
1295.64
1333.29
129.10
147.02
195.05
242.11
296.23
358.15
403.13
494.47
607.56
892.58
Relative Abundance
m/z
0
25
50
75
100
200
400
600
800
1000
1200

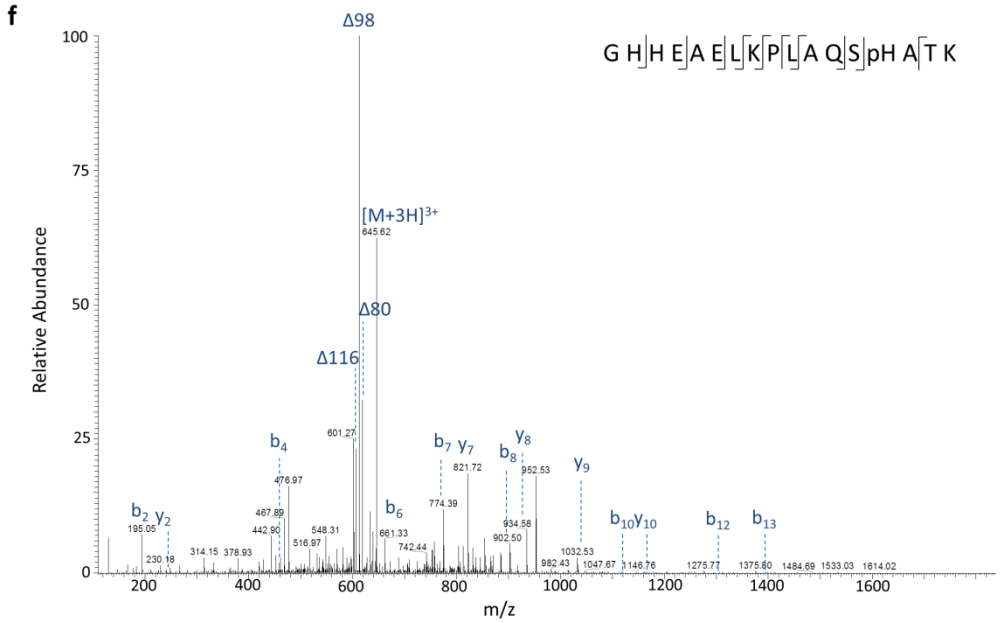
d


G pH H E A E L K P L A Q S H A T K
Δ98
y9
952.50
Δ116
Δ80
618.97
603.38
[M+3H]3+
y9 2+
476.86
y15 2+
y14 2+
b7/y8
y6
821.64
756.33
y7
742.40
830.64
934.47
y3
467.97
548.21
542.27
645.54
884.41
b8
y11
y12
y13
129.20
195.18
378.86
314.20
y2
248.26
671.37
1032.41
y10
1080.50
1186.61
1311.61
1376.69
1474.68
1532.87
1612.69
Relative Abundance
m/z
0
25
50
75
100
200
400
600
800
1000
1200
1400
1600

e


[M+2H]2+
968.54
G H pH E A E L K P L A Q S H A T K
[M+H]+
1935.80
[M+3H]3+
919.07
Relative Abundance
m/z

f


Δ98
G H H E A E L K P L A Q S pH A T K
[M+3H]3+
645.62
Δ80
Δ116
Relative Abundance
m/z

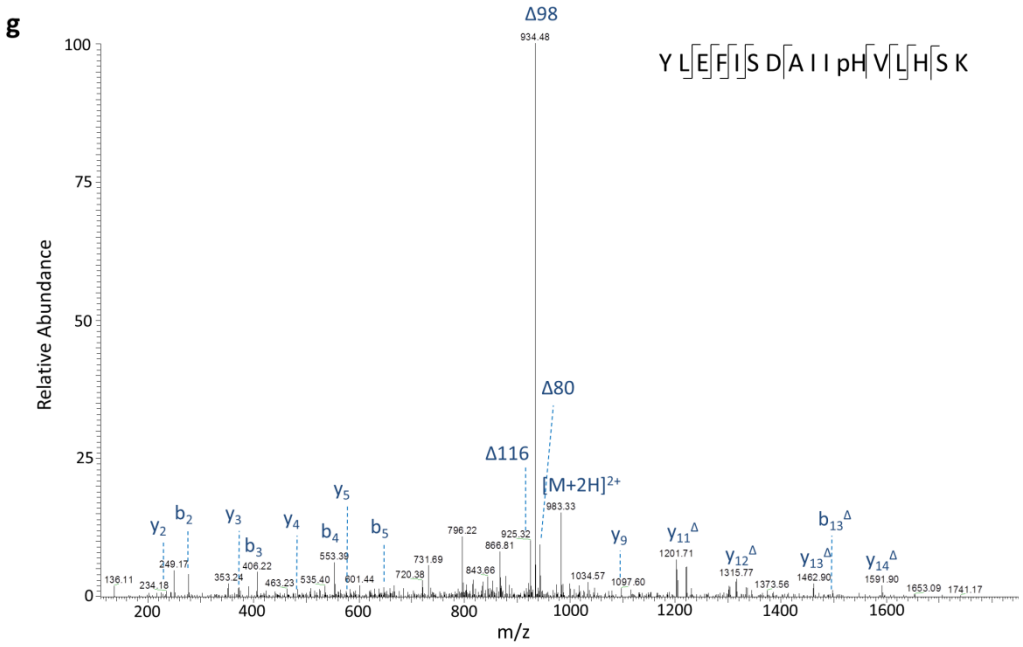


Figure 3.4 Product ion spectra of pHis-containing tryptic peptides of myoglobin generated by HCD or EThcD. a) doubly charged ion at *m/z* 845.42, HCD, pHis25; **b)** doubly charged ion at *m/z* 676.32, HCD, pHis37; **c)** doubly charged ion at *m/z* 730.41 HCD, pHis65; **d)** triply charged ion at *m/z* 645.31, HCD, pHis81; **e)** triply charged ion at *m/z* 645.31, EThcD, pHis82; **f)** triply charged ion at *m/z* 645.31, HCD, pHis94; **g)** doubly charged ion at *m/z* 983.50, HCD, pHis114. The sequence is displayed on the mass spectrum with the site of phosphorylation indicated as 'pH'.

3.2.2 Stability of phosphohistidine-containing peptides

Free pHis is known to be acid-labile[211, 212] and it is also reported to have increased stability at basic pH compared to Ser and Thr phosphopeptides[239]. There is however no published evidence regarding the pH stability of His-phosphorylated peptides. The pH stability of the myoglobin pHis-containing peptides was therefore assessed at 25 °C over two hours at pH 1, 4, 6 and 9. Acidic samples were neutralised at each time point by addition of ammonium hydroxide in an attempt to minimise additional de-phosphorylation which may have occurred between the time of collection and the point of analysis by mass spectrometry. The normalised intensity of two example pHis peptides is shown as a percentage compared to the amount at time = 0 (Figure 3.5). This data is an average of three technical replicates, where relative standard deviation between measurements is <20%. The results are representative of at least two independent experiments.

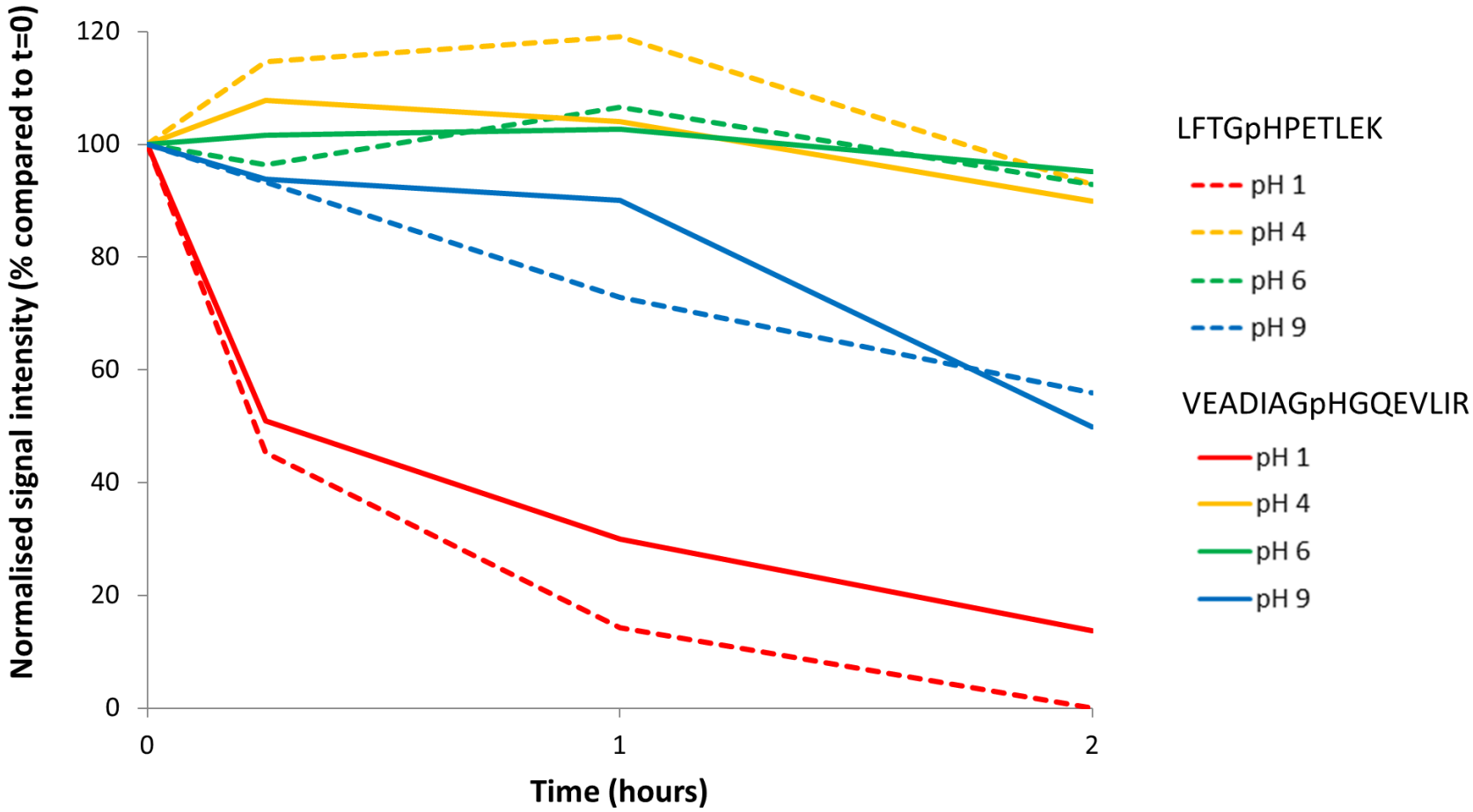


Figure 3.5 pH stability of pHis-containing tryptic peptides of myoglobin. Extracted ion chromatograms were used to determine the relative signal intensity of the pHis-containing peptides LFTGpHPETLEK at *m/z* 676.3 (dotted line) and VEADIAGpHGQEVLIR (solid line) at *m/z* 843.9 following incubation at pH 1, pH 4, pH 6 and pH 9. Signal intensity was normalised against a non His-containing peptide and is presented as a percentage compared to the amount at time = 0.

As expected the peptides show limited stability at pH 1. Both pHis peptides show rapid dephosphorylation, even after just 15 minutes, with the peptide VEADIAGpHGQEVLIR completely dephosphorylated by two hours, and the peptide LFTGpHPETLEK completely dephosphorylated by four hours. The $t_{1/2}$ of these phosphopeptides is ~15 minutes at pH 1, meaning that any enrichment technique that requires peptides to be under extreme acidic conditions even for a few minutes is likely to greatly impact the recovery of pHis peptides. The stability of these pHis peptides at pH 4 and pH 6 is much greater, with almost 100% of the peptides remaining even after two hours. At pH 9 these pHis peptides exhibit greater instability than would have been expected, with only around 60% of the phosphorylated peptide remaining after two hours. Given that previous hydrolysis studies have been performed only on the free pHis it is possible that stability is affected by other residues in the peptide. These results therefore provide suitable pH and time-scales upon which it should be possible to conduct sample enrichment and preparation techniques without incurring significant pHis dephosphorylation.

3.2.3 Titanium dioxide enrichment

TiO_2-based phosphopeptide enrichment is one of the most commonly used strategies in phosphoproteomics. The standard method employs acidic pH to improve specificity of binding[136], which given the previously discussed pH stability would not necessarily be suitable for analysis of pHis. Alternative approaches were therefore considered, including performing the binding step at non-acidic pH and replicating a previously reported method for the analysis of phosphoarginine (pArg) which is performed using lactic and acetic acid at pH 4[234]. A comparison of TiO_2 enrichment with these three sets of binding and elution conditions was performed; the buffer compositions used for binding, wash and elution steps are shown in Table 3.2 for each of the methods.

Enrichment efficiency was assessed by comparing the amount of each pHis-containing (and the corresponding non-phosphorylated) peptide in the un-enriched “start material”, the flow through and washes from the TiO_2 matrix, and the eluate (Table 3.3). Two replicates were performed and percentage recovery calculated for each peptide at each stage of the experiment. Recovery was calculated by determining the peak area for each peptide (phosphorylated and non-phosphorylated) in the flow through, washes and eluate, and expressing this as a percentage compared to the peak area of the same peptide in the un-enriched “start material”. The range of these calculated values across the five pHis peptides

is reported in Table 3.3. The large variability observed is due to differences in recovery between different peptides; typically variance between replicates for the same peptide was low.

Table 3.2 Conditions evaluated for TiO_2 enrichment of pHis (and other) phosphopeptides. Binding, sequential washing and sequential elution buffers for each of the three conditions (A, B and C) used to assess suitability of TiO_2 enrichment specifically for pHis-containing peptides are detailed. All % are (v/v).

	Binding	**Wash Steps**	**Elution Steps**
A	65% ACN, 2% TFA, saturated with glutamic acid, pH 2	1. 65% ACN, 0.5% TFA 2. 65% CAN, 0.1% TFA	1. 300 mM NH_3, 50% ACN 2. 500 mM NH_3, 60% ACN
B	65 mM NH_4OAc, 5% ACN, pH 7.5	1. 65% ACN, 0.5% TFA 2. 65% ACN, 0.1% TFA	1. 300 mM NH_3, 50% ACN 2. 500 mM NH_3, 60% ACN
C	3 M lactic acid, 60% ACN, 12.5% AcOH, pH 4	1. 2 M lactic acid, 75% ACN, 2% TFA 2. 2 M lactic acid, 75% ACN, 10% AcOH, pH 4 3. 80% ACN, 10% AcOH	1% NH_3, 30 mM $(NH_4)_3PO_4$

Table 3.3 Evaluation of TiO_2 enrichment of tryptic pHis peptides of myoglobin under three different sets of conditions. Range of values for binding and elution/recovery represents analysis from five pHis peptides for two replicate experiments. Binding, washing and elution buffers for experiments A, B and C are detailed in Table 3.2.

Enrichment conditions	**% Binding pHis peptides**	**% Elution pHis peptides**	**% Recovery non-phosphorylated His peptides**
A. Acidic (pH 2) binding	95 - 100	0	54 - 133
B. Neutral (pH 7.5) binding	98 - 100	0	7 - 164
C. Lactic/acetic acid (pH 4) binding	96 - 100	0	18 - 109

All TiO_2 binding conditions were compatible with efficient binding of pHis-containing peptides as demonstrated by very small amounts (maximum 5%) of any given pHis-containing peptide being detected in the flow through and washes. However, this did not translate into successful enrichment as none of the pHis-containing peptides of myoglobin were detected following elution from the TiO_2 material. It appears that under the acidic conditions of the binding /washing steps pHis peptides are dephosphorylated. This idea is supported by the fact that recovery of non-phosphorylated equivalents of the pHis peptides of myoglobin are greater than 100%. Extracted ion chromatograms (XICs) show that His-phosphorylated peptides are dephosphorylated during the enrichment process and therefore the amount of non-phosphorylated peptide recovered is greater than the amount in the "start material" (Figure 3.6).

Given the previous pH stability data this result is not surprising for the experiment conducted under the typical TiO_2 enrichment conditions (binding at pH 2) where rapid dephosphorylation would be expected. The pH stability data suggests that pHis peptides could be enriched at pH 4 and pH 7.5 without undergoing considerable dephosphorylation. However, these experiments still employ acidic wash steps which although were performed rapidly, appear to have significantly hindered the recovery of His-phosphorylated peptides.

In all cases enrichment of non-pHis-containing phosphopeptides was shown to be successful by the efficient recovery of pSer/pThr peptides of α-/β-casein. The average recovery of four representative phosphopeptides from α-/β-casein following TiO_2 enrichment was 67%.

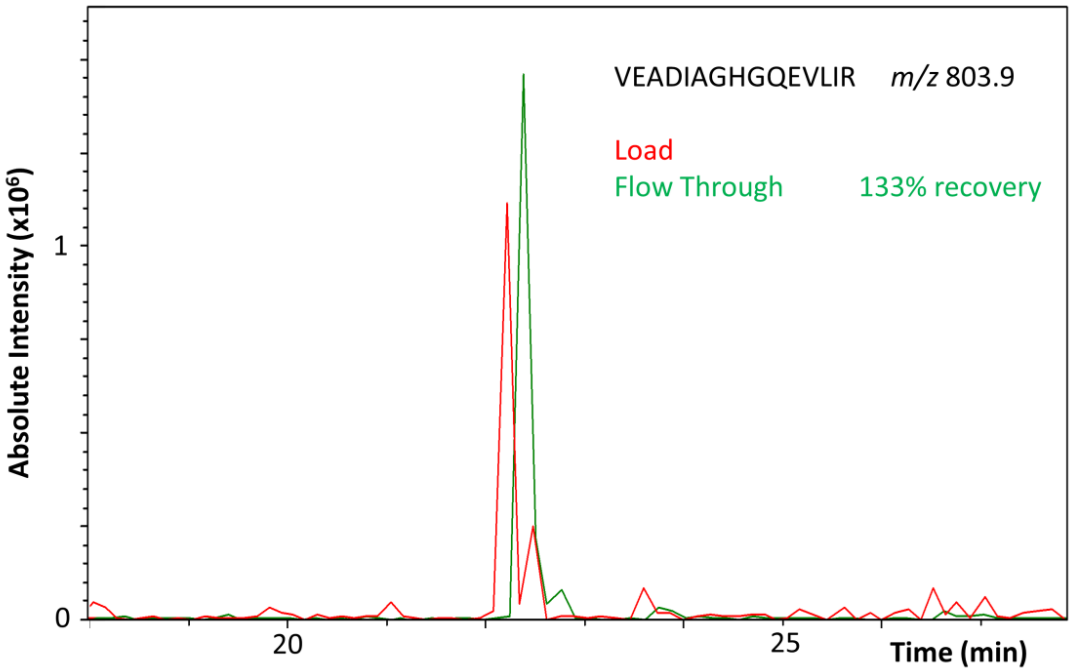


Figure 3.6 Representative XICs demonstrate dephosphorylation of pHis-containing peptides during TiO_2 enrichment. Overlay of XICs for the non-phosphorylated counterpart of the pHis-containing myoglobin peptide VEADIAGHGQEVLIR at *m/z* 803.9 in the load (red) and flow through (green) for TiO_2 enrichment performed under standard conditions. Greater than 100% recovery of the non-phosphorylated peptide indicates loss of the phosphate group from the corresponding pHis-containing peptide.

3.2.4. Hydroxyapatite enrichment

Following on from its application in chromatography columns to separate phosphoproteins, hydroxyapatite has been used as a solid support to enable phosphopeptide enrichment[150, 240]. The potential advantage of this technique for pHis analysis is that all binding, washing and elution steps are performed at non-acidic pH; binding to the solid support is performed at pH 7.2, and under standard conditions a phosphate salt at pH 7.8 competitively elutes phosphopeptides[155, 156]. Initial experiments were performed using these standard elution conditions, but poor recovery of both phosphorylated and non-phosphorylated His-containing peptides was observed (Table 3.4). In an attempt to improve specificity of phosphopeptide binding and recovery of pHis peptides the elution step was also performed at pH 6 and pH 7. Phosphopeptide enrichment efficiency was assessed by calculating the percentage recovery of phosphorylated and non-phosphorylated peptides in the flow though, washes and eluate compared to un-enriched "start material". Two replicates were performed and the values for % recovery displayed in Table 3.4 represent the range across five pHis peptides. As with the TiO_2 experiments the observed variability is due to differences in recovery between different peptides.

Table 3.4 Evaluation of hydroxyapatite enrichment of tryptic pHis peptides of myoglobin under three different sets of elution conditions. Range of values for binding and elution/recovery represents analysis from five pHis peptides for two replicate experiments.

Enrichment conditions (elution pH)	% binding pHis-containing peptides	% elution pHis-containing peptides	% recovery non-phosphorylated His-containing peptides
pH 7.8	83 - 96	0	4 - 18
pH 7.0	83 - 96	0	2 - 37
pH 6.0	83 - 100	0	1 - 35

Binding of pHis-containing peptides to the hydroxyapatite solid support was generally good, ranging from 83-100% across all experiments. However, none of the pHis-containing peptides from myoglobin were detected in the elution at any of the pH conditions trialled. Under the same conditions, pSer and pThr peptides from α-/β-casein were successfully enriched demonstrating that the column and conditions used were generally suitable for phosphopeptide enrichment. The average recovery of four representative α-/β-casein phosphopeptides was 43%. Unlike for the acidic TiO_2 experiments the lack of detection of

these pHis peptides was not suspected to be a result of dephosphorylation. Indeed, the recovery of non-phosphorylated His-containing peptides was less than 40%, indicating that both phosphorylated and non-phosphorylated His-containing peptides were not being eluted from the hydroxyapatite. The degree of peptide recovery was impacted very little by the pH of the elution buffer.

The results of the experiments performed here seem to demonstrate an exceptionally strong binding interaction between hydroxyapatite and these particular pHis peptides, and indeed their non-phosphorylated counterparts. Perhaps other residues of the peptides are binding and thus increasing the strength of the interaction, or His itself exhibits a particularly strong binding interaction with hydroxyapatite. Additionally, the average recovery of seven representative non-phosphorylated peptides from α-/β-casein is only 44% indicating that considerable non-specific binding could be causing poor overall recovery of peptides from hydroxyapatite. The results presented here for the attempted enrichment of pHis peptides of myoglobin using hydroxyapatite seem to suggest that it is not a suitable technique for analysis of pHis peptides.

3.2.5. Calcium phosphate precipitation

Another published method for phosphopeptide enrichment works by co-precipitation of phosphorylated peptides with calcium phosphate at basic pH[160, 161]. The non-acidic conditions, at least for the precipitation and wash steps, make this an appealing strategy for the analysis of pHis. The phosphopeptide recovery step is however performed under acidic conditions (5% TFA, ~pH 1) which may be expected to impact on the effectiveness of pHis-peptide recovery by this method. An attempt was made to rapidly re-solubilise the pellet with 0.1% TFA (~pH 3), in order to reduce potential dephosphorylation of pHis peptides. Re-solubilised samples were immediately loaded onto C18 desalting tips and desalted under neutral conditions.

The amount of phosphorylated and non-phosphorylated peptides in the flow through/washes (not precipitated) and the sample following re-solubilisation of the precipitate was compared to un-enriched “start material” to determine the percentage recovery of (phospho)peptides. Three replicates were performed and the values for % recovery displayed in Table 3.5 represent the range across five pHis peptides.

Table 3.5 Evaluation of calcium phosphate precipitation for enrichment of tryptic pHis peptides of myoglobin under two different sets of re-solubilisation conditions. Range of values for capture and recovery represents analysis from five pHis peptides for three replicate experiments.

Re-solubilisation conditions	% pHis- containing peptides captured in precipitate	% pHis- containing peptides recovered after re-solubilisation	% recovery (total) non-phosphorylated His-containing peptides
5% TFA	80 – 100	1 – 9	0 – 51
0.1% TFA	81 – 100	0 – 8	0 – 55

The calcium phosphate precipitation step captures at least 80% of the pHis-containing peptides, however, recovery of these peptides following re-solubilisation of the precipitate is extremely low (maximum recovery of 9% for pHis peptides). With 5% TFA this could be due to the expected dephosphorylation that would occur at pH 1. Whilst recovery of non-phosphorylated His-containing peptides didn't exceed 100% (as was demonstrated for dephosphorylation during TiO_2 enrichment) the non-phosphorylated His-containing peptides were detected in the sample following re-solubilisation. The recovery of pHis peptides at 0.1% TFA was potentially impacted by poor re-solubilisation at this pH; a pellet was still visible in the sample despite vigorous mixing, and this resulted in two of the peptides exhibiting 0% recovery in at least one replicate. This poor level of recovery was also seen for α-/β-casein peptides, whereby on average only 3% recovery was observed for four representative pSer/pThr peptides. Overall, the recovery of all peptides is lower than would be expected. A maximum of 55% of any non-phosphorylated His-containing peptide was recovered across the whole experiment and similarly recovery of α-/β-casein phosphopeptides (with 5% TFA) was only 48% on average. Sample losses through the course of the experiment, perhaps during the C18 desalting step, could account for this low recovery. The acidic re-solubilisation step has proven to be unsuitable for analysis of pHis peptides, with even a modest decrease in acidity to aid pHis peptide stability dramatically reducing enrichment efficiency.

3.2.6 Subtractive approaches for phosphohistidine peptide enrichment

Given the challenges associated with analysis of pHis-containing peptides using standard phosphopeptide enrichment strategies, an alternative approach was developed which would instead rely on indirect identification of pHis-peptides. A subtractive approach, as depicted in Figure 3.7, requires phosphopeptides to be bound to a solid support, such as TiO_2 or hydroxyapatite, whilst all non-phosphorylated peptides are washed away. Selective dephosphorylation of pHis is then performed and the recovered peptides are inferred as having previously been phosphorylated on His. Any remaining phosphopeptides (pSer/pThr) can be recovered using the standard elution conditions for each of the solid supports. The first step in establishing this approach was to determine suitable conditions for pHis peptide dephosphorylation. Previously presented results regarding the pH stability of pHis peptides of myoglobin indicate that an acidic treatment step could be used to elicit the required dephosphorylation.

As an alternative to acid, heat treatment may also be a viable option to selectively dephosphorylate pHis-peptides for the purposes of a subtractive workflow. Therefore, the stability of pHis-containing myoglobin peptides was assessed at elevated temperature. Myoglobin peptides were incubated in a neutral buffer at 80 °C and 95 °C for up to four hours. As for the pH stability studies, the normalised intensity of two example pHis peptides is shown as a percentage compared to the amount at time = 0 (Figure 3.8).

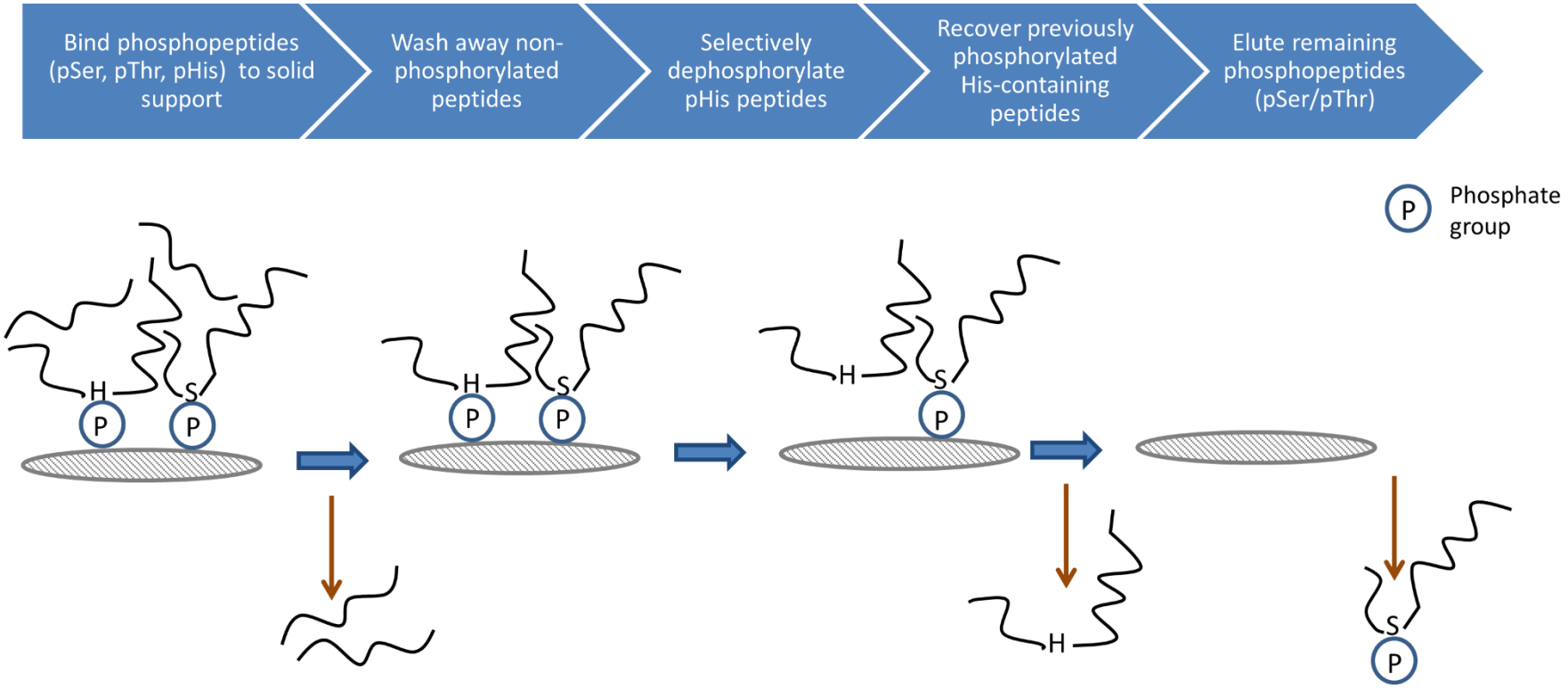


Figure 3.7 Workflow to show proposed subtractive strategy for pHis peptide analysis as an alternative to standard phosphopeptide enrichment workflows. Binding and wash steps should be performed at (near) neutral pH. Heat, acid and hydroxylamine treatment can be used to selectively dephosphorylate pHis-containing peptides with consideration given to compatibility with the proposed solid support.

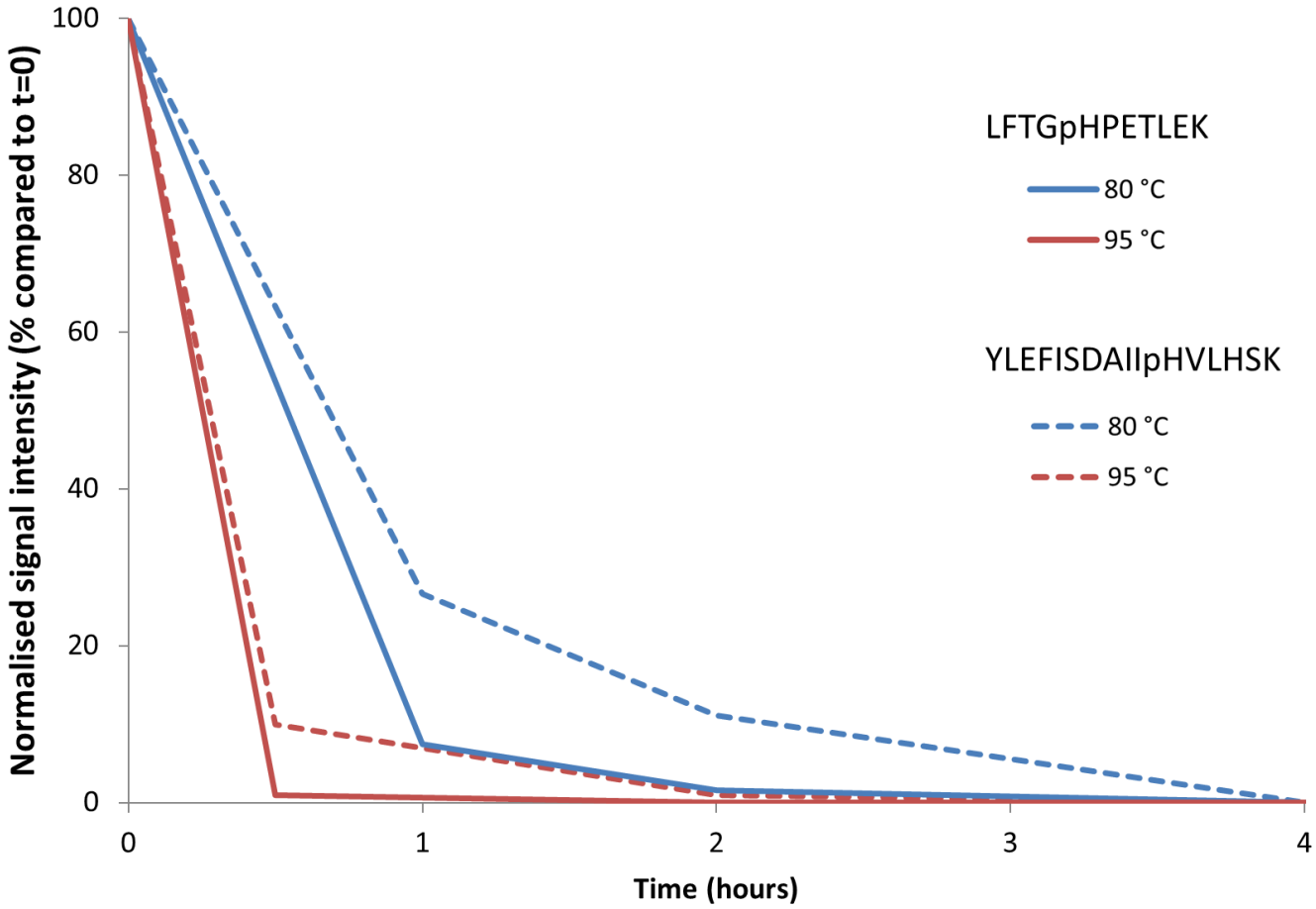


Figure 3.8 Temperature stability of pHis-containing tryptic peptides of myoglobin. XICs were used to determine signal intensity of the pHis-containing peptides YLEFISDAIIpHVLHSK at *m/z* 655.6 and LFTGpHPETLEK at *m/z* 676.3 following incubation at 80 °C and 95 °C. Signal intensity was normalised against a non His-containing peptide and is presented as a percentage compared to the amount at time = 0.

Rapid dephosphorylation was observed for both pHis peptides with a greater loss of the phosphate group observed at the higher temperature, as would be expected. At 95 °C the $t_{1/2}$ for both peptides is ~15 minutes, whilst at 80 °C the $t_{1/2}$ is 45-60 minutes. Based on these results it should be possible to use 95 °C heat treatment for the selective dephosphorylation of pHis within a subtractive workflow.

Another potential approach to elicit dephosphorylation of pHis-containing peptides is through the use of hydroxylamine, which has previously been shown to selectively dephosphorylate pHis[239]. The reduction in signal intensity of pHis-peptides of myoglobin following reaction for 1 hour with hydroxylamine is shown in Table 3.6. This data is from a single experiment. Although only four of the previously identified pHis peptides were observed in this experiment, presumably due to degradation of the pHis-myoglobin standard over time, this was deemed sufficient to assess the effects of hydroxylamine treatment. A representative XIC (Figure 3.9) for a single peptide pre-and post-hydroxylamine treatment indicates almost complete dephosphorylation of this peptide in solution.

Table 3.6 Reduction in pHis-containing peptide signal following reaction for 1 hour with hydroxylamine. XICs were used to determine the signal intensity for each of the pHis-containing peptides before and after the reaction, with the difference expressed as a percentage compared to the amount before the reaction.

Sequence	***m/z***	**% reduction in pHis peptide signal intensity**
LFTGpHPETLEK	676.3	93.6
pHGTVVLTALGGILK	729.9	91.8
VEADIAGpHGQEVLIR	843.9	95.1
YLEFISDAIIpHVLHSK	655.6	86.9

All pHis-containing peptides detected before the reaction exhibited at least an 86% reduction in signal intensity following treatment for 1 hour with hydroxylamine. Although the amount of dephosphorylation appears to be peptide dependent, with the amount of peptide remaining after the reaction ranging from 5-13%, this is nevertheless an effective method to dephosphorylate pHis peptides. Hydroxylamine is comparable to elevated temperature (95 °C) for pHis dephosphorylation, where ~5% of the pHis peptides were remaining after 1 hour. Both techniques perform slightly better than acidic treatment (pH 1) which after 1 hour exhibits a 70 – 85% reduction in pHis signal intensity.

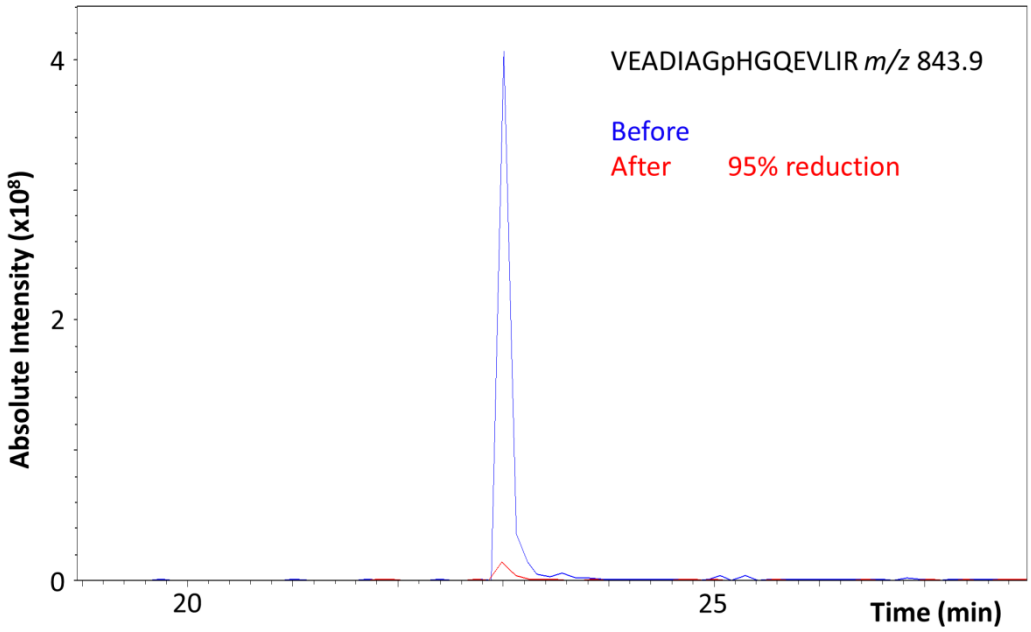


Figure 3.9 Representative XICs to demonstrate loss of phosphate following reaction of histidine phosphorylated myoglobin peptides for 1 hour with hydroxylamine. Overlay of XICs of pHis-containing peptide VEADIAGpHGQEVLIR at *m/z* 843.9 before (blue) and after (red) reaction with hydroxylamine shows 95% reduction in signal intensity.

Following on from this evaluation of pHis dephosphorylation, the subtractive approach was applied to both a TiO_2 and hydroxyapatite based strategy, according to the workflow previously depicted in Figure 3.7. Phosphopeptides were bound to the TiO_2 or hydroxyapatite solid support under (near) neutral conditions. For the TiO_2 experiment selective dephosphorylation was attempted using an acid treatment step, whilst for hydroxyapatite, which exhibits instability below pH 5.5, elevated temperature and hydroxylamine treatment were trialled. Percentage binding and recovery for pHis peptides and their non-phosphorylated equivalents was determined at each stage of the experiment. The range of these calculated values across the five pHis peptides is reported in Table 3.7.

Table 3.7 Recovery of tryptic pHis peptides of myoglobin by subtractive approach using TiO_2 or hydroxyapatite. Binding and recovery of pHis peptides (and non-phosphorylated equivalents) following a subtractive strategy for enrichment using acid to elicit selective dephosphorylation of pHis peptides bound to TiO_2 and either heat or hydroxylamine for pHis peptides bound to hydroxyapatite. Range of values represents analysis of five pHis peptides.

Solid support & dephosphorylation conditions	**% binding pHis-containing peptides**	**% binding non-phosphorylated His-containing peptides**	**% recovery non-phosphorylated His-containing peptides after treatment**
TiO_2: Acid (pH 1)	89 - 100	0 - 94	0
Hydroxyapatite: Heat (95 °C, 45 mins)	78 – 93	18 – 68	0 – 2
Hydroxyapatite: Hydroxylamine (1 hour)	85 – 96	0 – 71	0 – 2

Under these conditions binding of pHis peptides to the TiO_2 solid support was efficient (89-100%) but a number of other problems were highlighted with the proposed method. Firstly, there was considerable binding of some of the non-phosphorylated His-containing peptides to the TiO_2 solid support (up to 94%) when this was performed at neutral pH. Binding of non-phosphorylated peptides is undesirable as they could interfere with subsequent identification of His phosphorylated peptides. Secondly, there was no recovery of previously phosphorylated His-containing peptides following the acid treatment step, with a large portion of these peptides remaining unaccounted for in any stage of the experiment. The reasons for this are currently unexplained. Finally, following the acid treatment a few non-phosphorylated peptides of α-/β-casein are observed, suggesting this step can remove peptides from the TiO_2 solid support. Elution of these non-phosphorylated peptides is detrimental in this type of subtractive strategy as such peptides could mistakenly be identified as having previously been phosphorylated on His in a more complex sample where sites of phosphorylation are unknown.

Both hydroxyapatite based experiments demonstrate reasonable binding of pHis-containing peptides to the hydroxyapatite resin, as was previously observed with this solid support, therefore highlighting that the binding stage of this experiment is robust. However, despite both elevated temperature and reaction with hydroxylamine proving effective for almost complete dephosphorylation of pHis peptides in solution (Figures 3.8 and 3.9), these treatments did not facilitate recovery of peptides bound to hydroxyapatite. Neither incubation at 95 °C for 45 minutes or with hydroxylamine for 1 hour resulted in recovery of His-containing peptides from hydroxyapatite. An additional issue with this strategy is the incomplete removal of non-phosphorylated peptides from the resin prior to the dephosphorylation step; as little as 30% is recovered in flow through/washes in some cases. The apparently strong binding of non-phosphorylated peptides to hydroxyapatite resin in these experiments is perhaps not unexpected given the poor recovery of peptides previously observed for the hydroxyapatite enrichment strategy (Table 3.4). Binding of non-phosphorylated peptides at this stage could hinder interpretation of results from more complex experiments with unknown phosphosites whereby pHis peptides would be inferred by their lack of a phosphate group.

3.2.7 Derivatisation of histidine residues

An additional strategy for indirect identification of pHis residues was attempted based on chemical derivatisation of non-phosphorylated His residues. The general workflow for such a strategy is detailed in Figure 3.10. With this approach His residues would first be chemically modified then pHis peptides could either be subjected to selective dephosphorylation or standard sample preparation techniques which relied on acid could be used without the loss of the phosphate group hindering analysis. The absence of a phosphate group (alongside the lack of the chemical derivatisation) would be indicative of His phosphorylation.

DEPC and HNE have both been previously shown to be capable of modifying His[235, 236], according to the reactions displayed in Figure 3.11. Tryptic peptides of the phosphorylated myoglobin standard were reacted with both DEPC and HNE to observe the extent of modification of unphosphorylated His. Both reagents were shown to be capable of modifying His-containing peptides. The efficiency of the reaction is displayed as a percentage reduction in His-containing peptide signal intensity (Table 3.8), as this can be compared before and after the reaction. Following the reactions chemically modified His-containing peptides were identified by mass spectrometry by addition of 72 Da (DEPC) or 156 Da (HNE) to the non-modified peptide mass, thus indicating that the observed reduction in His-containing peptide signal is a result of the expected modification having occurred. Peptides without a His residue remained unmodified by these reactions, as determined by comparable signal intensity before and after the reaction.

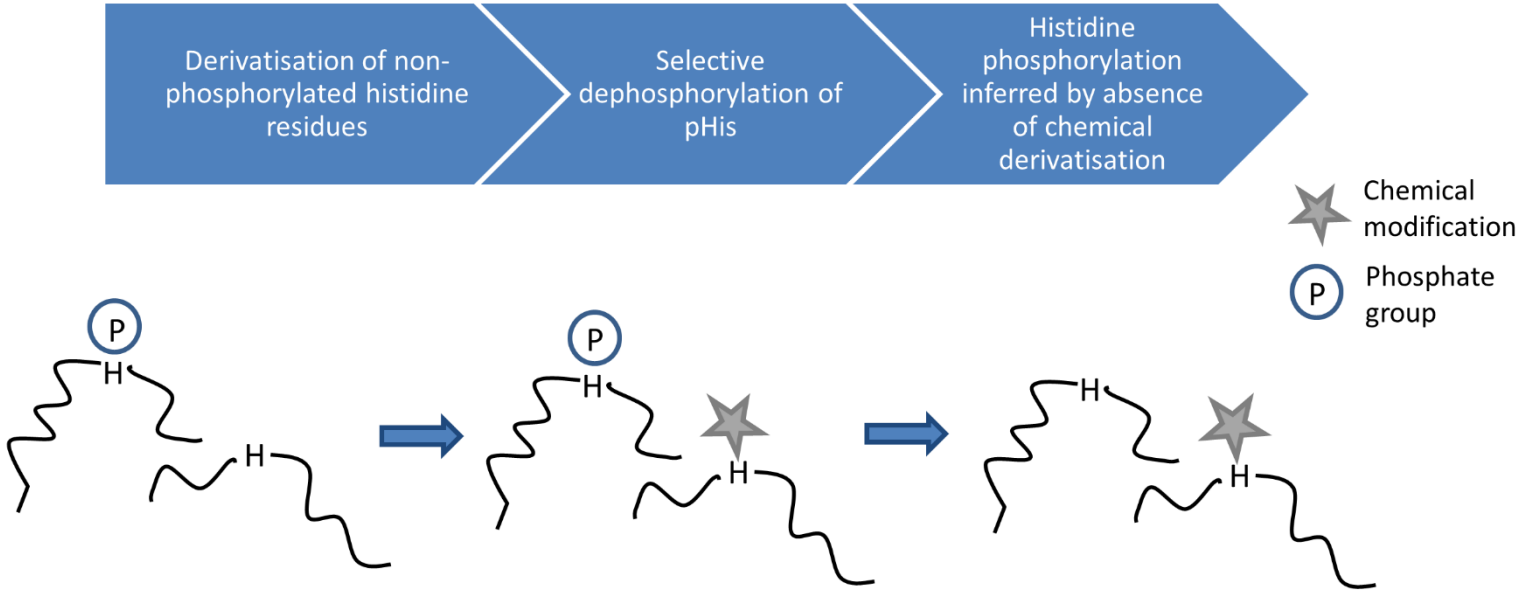


Figure 3.10 Workflow to show subtractive strategy based on chemical derivatisation of non-phosphorylated histidine residues. Dephosphorylation of pHis peptides following chemical derivatisation of non-phosphorylated His residues allows His phosphorylation to be inferred by lack of phosphate group or chemical modification.

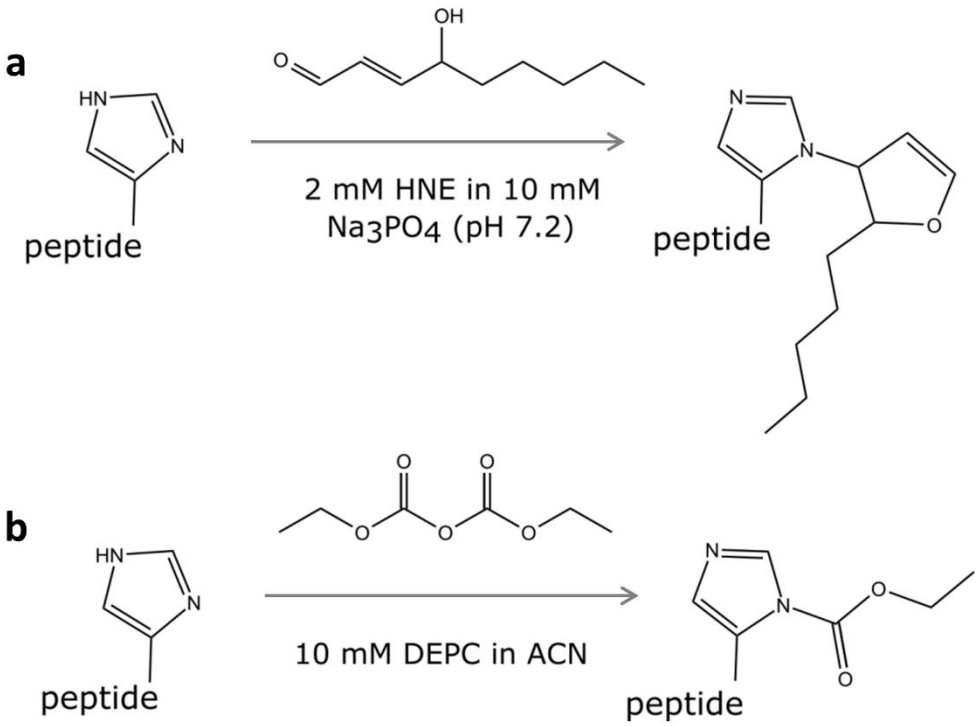


Figure 3.11 Reaction scheme for chemical modification of histidine residues. Non-phosphorylated His residues can be modified by reaction with **a)** 4-hydroxynonenal (HNE) or **b)** diethylpyrocarbonate (DEPC)

Table 3.8 Reduction in signal intensity of histidine-containing peptides following reaction with 4-hydroxynonenal (HNE) and diethylpyrocarbonate (DEPC). XICs were used to determine the signal intensity for each of the His-containing peptides before and after reaction, with the difference expressed as a percentage compared to the amount before the reaction.

Peptide	***m/z***	**% reduction in signal intensity**	
		HNE	**DEPC**
LFTGHPETLEK	636.3	47.3	68.8
HPGDFGADAQGAMTK	751.8	36.9	69.6
VEADIAGHGQEVLIR	803.9	61.5	42.6
HGTVVLTALGGILK	689.9	99.4	3.8

For both reactions the degree of modification of His-containing peptides is variable and incomplete. With HNE one peptide is almost completely modified (HGTVVLTALGGILK; 99%) although for others there is as little as 37% modification. For reaction with DEPC the degree of modification ranges from 4-70%, with this reaction performing better than HNE for some peptides but worse for others. For these reactions to be successfully incorporated into a subtractive strategy for pHis peptide identification, the degree of modification would need to approach 100% for all His-containing peptides to allow conclusions to be drawn that the now free His-peptides were previously phosphorylated. Further experiments may have revealed reaction conditions which facilitated complete reaction of all His residues, but this was not investigated any further. As it stands, chemical derivatisation of His-containing peptides with these reagents appears not to be suitable for use as part of a subtractive strategy for pHis peptide identification.

3.2.8 Immunoprecipitation

One of the most recent developments in the field of pHis analysis is the production of monoclonal antibodies for the 1- and 3- isomers of pHis[227]. These have been shown to be capable of enriching pHis proteins from cell lysate via an immunoprecipitation approach, although the published data does not give details about the His residues of these proteins that are phosphorylated. To determine effectiveness of these antibodies for pHis peptide analysis, an immunoprecipitation experiment was conducted using the 1- and 3-pHis antibodies to enrich for pHis-containing proteins in a U2OS lysate, which were subsequently digested and analysed by LC-MS/MS using the Orbitrap Fusion mass

spectrometer. Data was searched using MASCOT, with a peptide FDR of 1% and two peptides required per protein for identification.

A total of 1002 proteins were identified following immunoprecipitation with the 3-pHis antibody, compared to 837 with the 1-pHis antibody. There are 684 proteins common to both IP experiments (Figure 3.12 a). Many of the proteins identifications shared with both samples are likely to be a result of non-specific interactions, resulting in a background of common contaminant proteins binding to the affinity resin[241, 242]. The data generated in these experiments was compared with previously published data for pHis immunoprecipitation (Figure 3.12 b, c and full list of overlapping proteins in Appendix). Proteins from the Fuhs *et al.* list were filtered to only include those which showed a greater than 2-fold difference compared to the control, with the aim of removing proteins that had bound non-specifically from this list. Overlap of protein IDs between this experiment and the previously published list was 24% for 3-pHis proteins and 28% for 1-pHis proteins. However, there were no pHis-containing peptides identified in either the 1- or 3-pHis immunoprecipitation experiments, and indeed very few phosphorylated peptides were identified in either sample. Without detection and identification of pHis peptides there is no way to validate, using only this type of enrichment, whether any of the identified proteins were actually phosphorylated on His. Despite enriching for pHis proteins, there are still a vast number of non-phosphorylated peptides present in the sample after tryptic digestion which will potentially result in the same ion-suppression effects evident in non-enriched samples and stoichiometry is still likely to be low. In order for this immunoprecipitation approach to be used for identification of pHis-containing peptides an additional stage of enrichment may be required at the peptide level following enrichment of pHis proteins.

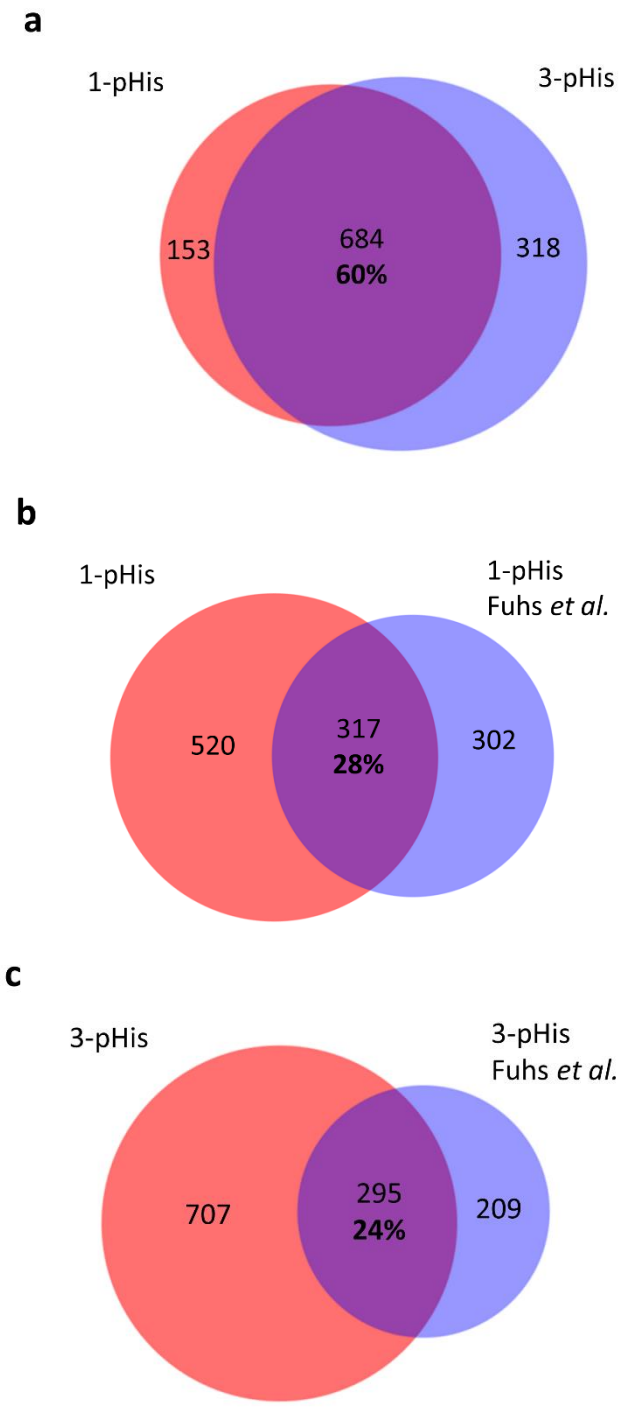


Figure 3.12 Immunoprecipitation with 1- and 3-pHis antibodies a) Overlap in protein IDs between 1-and 3-pHis IP experiments **b)** Overlap in protein IDs between this experiment and published IP data for 1-pHis **c)** Overlap in protein IDs between this experiment and published IP data for 3-pHis (List of overlapping protein IDs can be found in the Appendix)

3.3. Conclusions

His-phosphorylated peptides have been generated by chemical phosphorylation of myoglobin with potassium phosphoramidate and subsequent tryptic digestion. MS/MS analysis was used to identify five tryptic pHis-containing peptides, and the sites of His phosphorylation were determined. As expected based on previously published studies regarding pH stability of the free pHis amino acid[211, 212], these pHis-containing peptides exhibit considerable instability at acidic pH and high temperatures, which poses a number of challenges for their analysis by typical (phospho)proteomics strategies. It was also observed that these pHis peptides were less stable at pH 9 than was previously expected.

None of the methods presented in this chapter have proven successful for the analysis of pHis-containing peptides. Attempts to enrich with either TiO_2 or hydroxyapatite resulted in no recovery of pHis-containing peptides. In the case of TiO_2, the primary reason for the failure of this method was the dephosphorylation of pHis over the course of the experiment, whilst the hydroxyapatite method appears to display considerable non-specific binding. Both methods were shown to enrich pSer/pThr-containing peptides meaning the aforementioned issues are specific to pHis. Calcium phosphate precipitation also proved unsuccessful for pHis peptide enrichment due to dephosphorylation of pHis peptides during acidic resolubilisation of the precipitation, and was further hindered by poor peptide recovery.

Additional methods trialled were based around a subtractive approach, using TiO_2 or hydroxyapatite for binding followed by selective dephosphorylation of pHis-containing peptides by reduction in pH, elevated temperature or hydroxylamine treatment. Again, problems associated with these methods, based on extremely poor recovery meant that they could not be used for pHis peptide analysis. A similar subtractive approach was tested based on the chemical derivatisation of non-phosphorylated His residues whereby His phosphorylation would be inferred by the absence of the chemical modification (or phosphorylation) following standard phosphoproteomics strategies. Neither DEPC nor HNE were capable of consistent and complete derivatisation of non-phosphorylated His residues which would be essential for such an approach to be feasible.

An attempt was also made to identify pHis-containing peptides and proteins via an immunoprecipitation approach using monoclonal 1- and 3-pHis antibodies. Despite a large number of proteins being identified, approximately 25% of which match with previously published data, there were no pHis-containing peptides identified in the dataset, meaning that the identity of these proteins as His phosphorylated can only be inferred.

Overall, it would be preferable for any strategy for pHis analysis to be based upon the direct identification of pHis-containing peptides, therefore allowing the site of phosphorylation to be localized to a specific residue within the peptide. His is not the only labile phosphorylation suspected to be present in mammalian systems; Lys, Arg, Asp, Glu and Cys are also known to be phosphorylated in other systems, although there is little to no evidence currently for these modifications in higher eukaryotes, arguably largely because of difficulty in their identification. An indirect approach to phosphopeptide identification would ignore modification of other acid-labile residues, and peptides with other phosphorylated residues may be missed, or indeed mis-assigned as pHis. Ideally, an analytical strategy developed for characterisation of pHis would also be suitable for identification of other acid-labile phosphorylated residues, allowing for an unbiased approach to phosphopeptide analysis.

Chapter 4. Results II: Development of a strong anion exchange approach for analysis of phosphohistidine

4.1 Introduction

Following on from the unsuccessful attempts to adapt existing phosphopeptide strategies for enrichment and subsequent characterisation of pHis (*Chapter 3. Results I*) an alternative approach was sought. This chapter describes the development of a strong anion exchange (SAX) fractionation method suitable for analysis of pHis-containing peptides. Initial evaluation of the method with a simple peptide mixture is followed by validation for its suitability for analysis of a complex cell lysate.

A fractionation strategy is different to enrichment in that all peptides are recovered and analysed, compared to enrichment whereby the aim is to remove non-phosphorylated peptides, as far as is possible, from the analysis. Fractionation is routinely used in proteomics studies to increase the depth of proteome coverage achieved, with strong cation exchange (SCX) and high pH reverse-phase chromatography being particularly attractive for whole proteome analysis[84, 169, 173]. Phosphoproteome analysis has also been conducted using SCX[243] although more recent developments have seen SCX coupled to phosphopeptide enrichment techniques such as IMAC[92, 127], or combined with an additional fractionation step such as anion exchange[177]. SAX has proven to be particularly effective for analysis of phosphopeptides, either as a strand alone fractionation strategy or coupled with enrichment steps or another dimension of separation such as high pH reversed phase chromatography to improve coverage[178-180, 244, 245].

All ion exchange, including SAX, enables separation of peptides based on their net accessible charge, which is dependent on the residues a given peptides possess and the pH at which the separation is conducted. A decrease in pH results in increased protonation of amino acid residues within the peptide whose side-chains possess an ionisable group, thereby reducing the ion-ion interaction these peptides have with a positively charged stationary phase and/or introducing a repulsive effect. Therefore, a decreasing pH gradient can be used to perform anion exchange separation of peptides. Alternatively, anion exchange can be conducted at a constant pH with elution brought about by an increasing salt gradient; peptides with a stronger interaction with the anion exchange column due to greater net negative charge require a higher concentration of salt to be eluted. In terms of

phosphopeptides, the negatively charged phosphate group will increase the interaction between these peptides and the SAX column, meaning they will elute later in the gradient, regardless of whether a pH or salt gradient is used. A recently reported method by Alpert et al[180] demonstrates that strong anion exchange based fractionation of phosphopeptides can be achieved using triethylammonium phosphate, with binding at high pH and elution by pH gradient. The work described in this chapter seeks to develop the SAX methodology to allow fractionation of pHis peptides using a salt gradient at non-acidic pH.

4.2 Results

4.2.1. Optimal pH for SAX fractionation

To help maintain pHis peptide stability, the SAX fractionation strategy must be conducted at non-acidic pH, using a salt gradient for peptide elution. The optimal conditions for fractionation of phosphopeptides are such that separation of phosphorylated and non-phosphorylated peptides is maximised. Negatively charged phosphate groups ensure increased retention of phosphopeptides, which would be expected to result in SAX fractions at the end of the gradient containing a higher percentage of phosphopeptides compared to those at the beginning where the salt concentration is low. The pKa values of a phosphate group and amino acid side chains in peptides[246] are given in Table 4.1.

Table 4.1 pKa values of phosphate group and amino acid side chains. Phosphate has three pKa values as shown. Amino acid side chain pKa values measured in peptides[246]

	pKa
Phosphate	2.2, 7.2, 12.4
Asp	3.9
Glu	4.3
His	6.5
Cys	8.6
Tyr	9.8
Lys	10.4
Arg	12.3
C terminus	3.7
N terminus	8.0

Above pH 2 all phosphate groups will have one negative charge, whilst increasing the pH above ~7 will result in phosphopeptides gaining an additional charge and therefore exhibiting a stronger interaction with the SAX column. Above pH ~4 all Glu/Asp residues will be deprotonated, thus there is likely to be some interference in later fractions from peptides with multiple acidic residues when conducting SAX at non-acidic pH. Finally, given that histidine has a pKa of 6.5, non-phosphorylated His residues will carry a positive charge below pH 6.5, resulting in peptides containing His eluting earlier in the gradient when SAX is conducted at lower pH. The difference in charge between protonated and phosphorylated His residues will allow for a greater separation between these peptides and their phosphorylated equivalents than would be observed for other pairs of

phosphorylated (pSer/pThr) and non-phosphorylated peptides, where the non-phosphorylated residue is uncharged.

Initial method development and optimisation was conducted using a standard mixture of α-/β-casein and myoglobin tryptic peptides, containing non-phosphorylated, pSer/pThr and pHis peptides. SAX fractionation was tested at pH 6, 6.8 and 8. Fractions were collected every minute and pooled to give 16 fractions overall, which were subsequently analysed by LC-MS/MS. An equal volume of each fraction was loaded onto the trap column of the LC-MS system at neutral pH, with the LC column then developed under standard conditions. The data was processed using PEAKS in order to perform label free quantification of each peptide and thus assess efficiency of peptide separation and recovery. The overall phosphopeptide recovery in each fraction was determined by summing all reported signal intensity for a given fraction, then calculating the percentage of this attributable to either phosphorylated or non-phosphorylated myoglobin and α-/β-casein peptides (Figure 4.1).

Under all pH conditions it was observed that the initial fractions consist of predominantly non-phosphorylated peptides, whilst the latter fractions show up to 100% of signal attributable to phosphorylated peptides. The SAX fractionation approach separates phosphorylated peptides from non-phosphorylated peptides using a salt gradient elution, and results in a number of fractions towards the latter end of the column development that are considerably enriched for phosphopeptides. At pH 6.8 there is the greatest number of fractions containing more than 80% phosphopeptide signal; 7 fractions have phosphopeptide signal >80% at pH 6.8, compared to 3 and 4 fractions at pH 6 and 8 respectively.

All three analyses show fractions towards the middle of the gradient (fractions ~10-13) that appear to have a higher percentage of non-phosphorylated peptides than might be expected based on the otherwise observed trend of decreasing non-phosphorylated peptide signal as the gradient progresses. This apparent discrepancy can be explained as an artefact of the data processing strategy. Signal from one non-phosphorylated peptide in particular may dominate the fraction despite the total number of non-phosphorylated peptides being low. For example, the peptide FQSEEQQQTEDELQDK from β-casein generates the most intense ion signal by far in fraction 12 at pH 6, resulting in almost 80% of the total signal attributed to non-phosphorylated peptides. However, the percentage of

non-phosphorylated peptides identified in this fraction in terms of unique number of peptides is only ~50% (12 out of 22 peptides are non-phosphorylated). Despite this issue, representing phosphopeptide enrichment in terms of signal intensity is overall preferable for a simple mixture, as the observed enrichment is not diminished by detection of small amounts of non-phosphorylated peptides in later fractions, potentially due to carry-over on the LC column. Overall, using peptide signal intensity is appropriate for comparison between conditions as is required for this experiment.

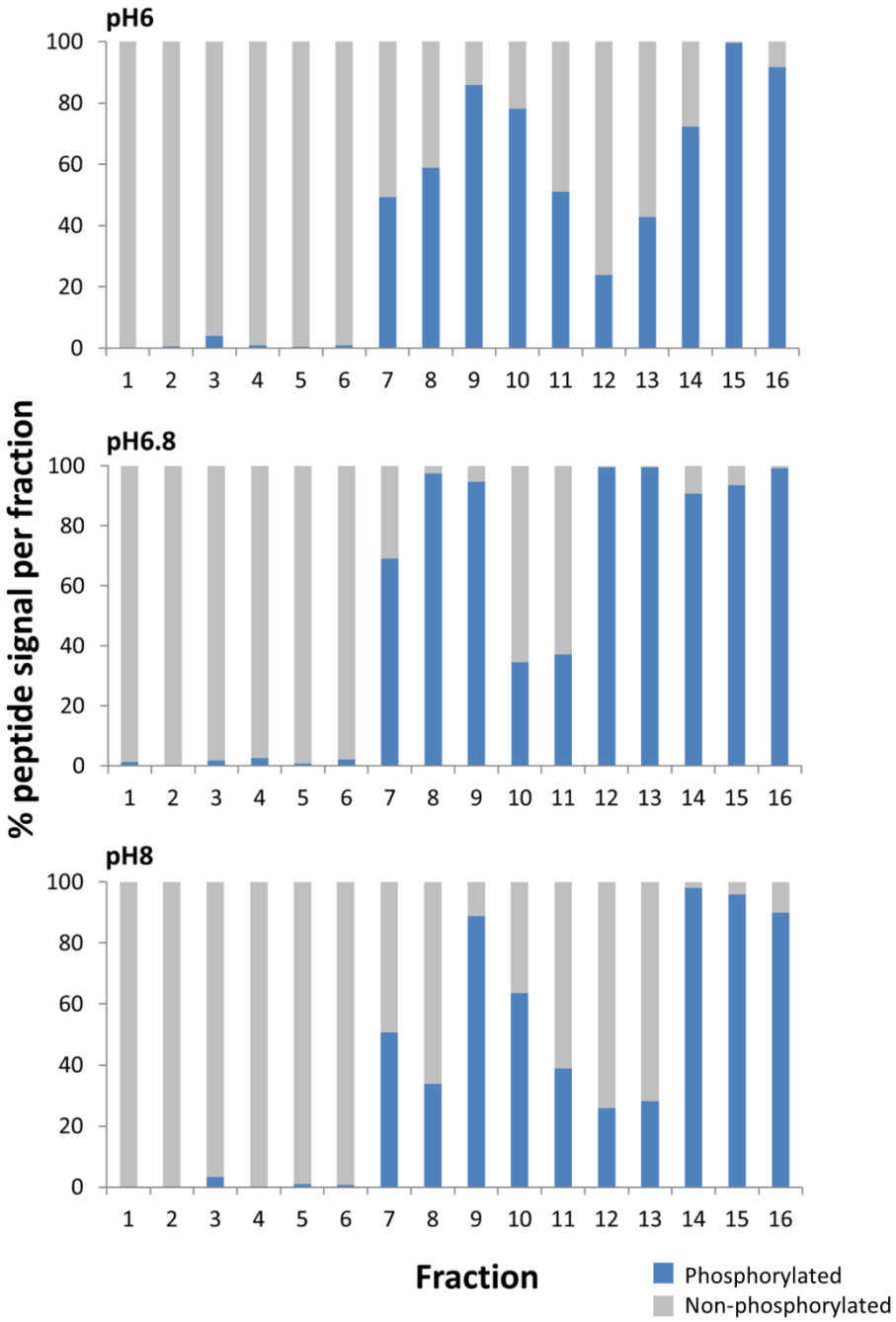


Figure 4.1. Separation of phosphorylated and non-phosphorylated peptides is achieved by strong anion exchange (SAX) chromatography. Percentage of the total signal intensity attributed to phosphopeptides (blue) and non-phosphorylated peptides (grey) in each SAX fraction at pH 6, 6.8 and 8, following LC-MS/MS analysis and label free quantification using PEAKS, with later fractions consisting of up to 100% phosphopeptides.

The label free quantification data was also used to determine the fractions in which pHis peptides of myoglobin (and their non-phosphorylated equivalents) were observed. For any given peptide, the total signal intensity across all fractions was summed, and the amount in each individual fraction then displayed as a percentage of this total amount. The results for each pair of the pHis and equivalent non-phosphorylated peptide are compared at different pH values (Figure 4.2).

At pH 8 (green lines), only two of the five pHis peptides (VEADIAGHGQEVLIR and LFTGHPETLEK, Figure 4.2 a and b) exhibit separation from their non-phosphorylated equivalent. The other three peptides (HGTVVLTALGGILK, GHHEAELKPLAQSHATK and YLEFISDAIIHVLHSK, Figure 4.2 c, d and e) elute in the same fraction as their non-phosphorylated equivalents as shown by the overlapping solid and dashed lines. In general, non-phosphorylated peptides elute in later fractions at pH 8 compared to at pH 6.8 and 6. Later elution of non-phosphorylated His-containing peptides at higher pH can be explained by the fact that the imidazole side chain of histidine (pKa 6.5) will be deprotonated at pH 8, yielding an uncharged residue, and resulting in reduction of the repulsive interaction with the SAX column that is observed when the histidine side chain is positively charged. The effects of this repulsive interaction between positively charged histidine and the SAX column at lower pH is particularly apparent for peptides HGTVVLTALGGILK and GHHEAELKPLAQSHATK which are observed in fraction 1 at pH 6 and pH 6.8.

The peptides VEADIAGHGQEVLIR and YLEFISDAIIHVLHSK show similar behaviour under all three pH conditions, with the pHis and non-pHis versions of YLEFISDAIIHVLHSK eluting in the same fractions and VEADIAGHGQEVLIR in adjacent fractions regardless of pH. As pH increases from pH 6 to pH 8, increased deprotonation of the phosphate group should result in greater retention of a given phosphorylated peptide and hence greater separation from the non-phosphorylated equivalent. In the case of these peptides other residues may be suppressing the effects of additional deprotonation, resulting in similar retention by the SAX column at pH 6, 6.8 and 8. However, for HGTVVLTALGGILK, GHHEAELKPLAQSHATK and LFTGHPETLEK differential separation is observed. Although maximum separation is observed at pH 6 (red lines) for the LFTGHPETLEK peptide, separation is maximised for the other two peptides at pH 6.8 (blue lines). Additionally, at pH 6 the non-phosphorylated HGTVVLTALGGILK peptide (and to a lesser extent the GHHEAELKPLAQSHATK peptide) elute in two different fractions (HGTVVLTALGGILK in fractions 1 and 4 and GHHEAELKPLAQSHATK

in fractions 1 and 6) as shown by a second peak for the dashed lines, which could potentially indicate on-column dephosphorylation of the pHis peptide. Taking in to account peptide stability and separation of phosphorylated and non-phosphorylated peptides, pH 6.8 appears to be optimal for separation of pHis peptides from their non-phosphorylated equivalents under conditions of pHis stability.

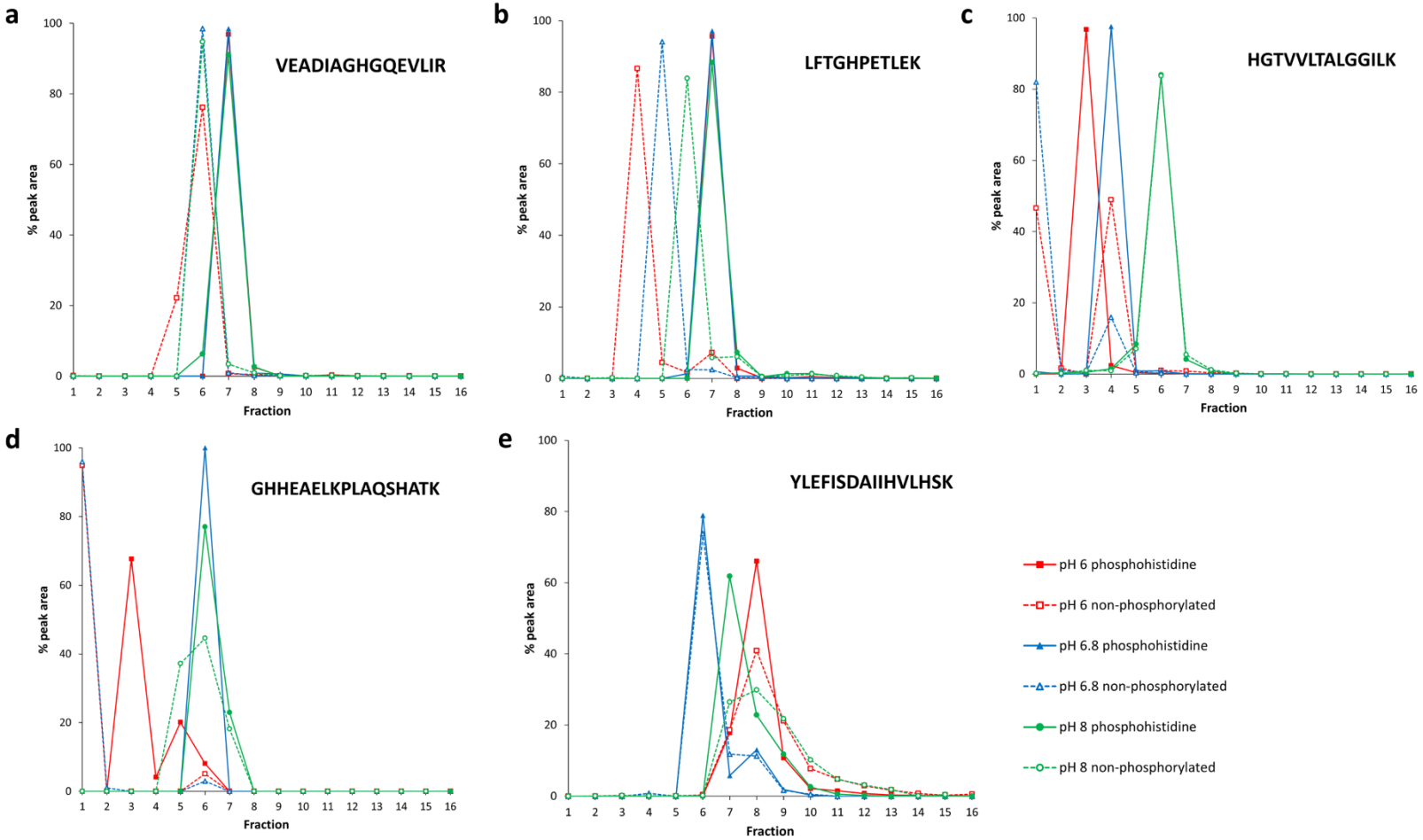


Figure 4.2 Differential elution of myoglobin pHis peptides and their non-phosphorylated equivalents at different pH values. Label free quantification of pHis-containing peptides (solid lines) and their non-phosphorylated equivalents (dashed lines) in individual SAX fractions, expressed as a percentage of the total ion current across all fractions, for separation at pH 6 (red square), pH 6.8 (blue triangle) and pH 8 (green circle)

A similar assessment of peptide separation under the three pH conditions was carried out for the peptides of α-/β-casein. The amount of phosphorylated and non-phosphorylated peptide in each fraction is displayed as a percentage of the total amount across all fractions for four representative peptides (Figure 4.3).

The phosphorylated peptides (YKVPQLEIVPNSAEER, VPQLEIVPNSAEER, TVDMESTEVFTK and FSSEEQQQTEDELQDK, Figure 4.3 a-d) and their non-phosphorylated counterparts elute in separate fractions under all pH conditions. A difference of at least two fractions in terms of where the majority of the phosphorylated and non-phosphorylated version of the peptide elutes is observed, as indicated by the non-overlapping dashed and solid lines of the same colour. The observation of non-phosphorylated FQSEEQQQTEDELQDK (Figure 4.3 d) in fractions 10 and 12 highlights the fact that peptides containing many acidic residues will elute in the later fractions along with phosphorylated peptides. Above pH 6 both Asp and Glu will be deprotonated (pKa 3.9 and 4.3 respectively) resulting in greater retention of peptides containing a high number of these residues compared to those without. The retention of acidic peptides could be minimised by lowering the pH, but this would have a detrimental effect on pHis peptide stability, hence their presence in the later fractions must be tolerated. As sample complexity increases, interference from acidic peptides may become more apparent, but this is not expected to have an overall negative impact on analysis of pHis peptides.

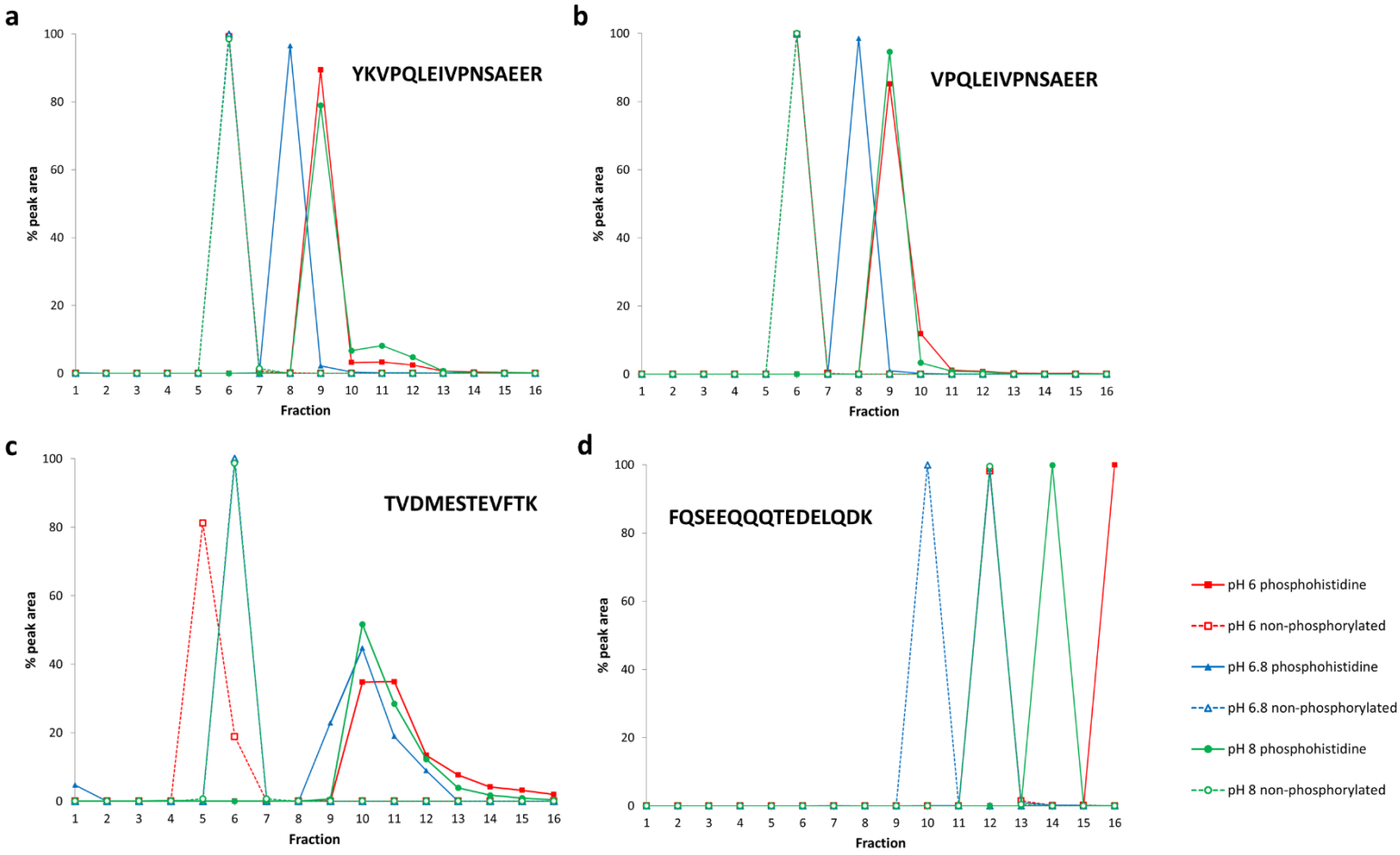


Figure 4.3 Differential elution of phosphorylated and non-phosphorylated peptides of α-/β-casein at different pH. Label free quantification of phosphorylated peptides (solid lines) and their non-phosphorylated equivalents (dashed lines) in individual SAX fractions, expressed as a percentage of the total ion current across all fractions, for separation at pH 6 (red square), pH 6.8 (blue triangle) and pH 8 (green circle)

4.2.2 Comparison of SAX with high pH reversed-phase fractionation

High pH RP fractionation has been utilised for phosphoproteomics analysis of complex cell lysates, typically coupled to an additional dimension of separation or combined with a phosphopeptide enrichment step, such as titanium dioxide or IMAC[174, 245, 247]. A comparison of high pH RP fractionation and SAX was conducted in order to establish whether the phosphopeptide enriched fractions obtained following SAX were purely a result of sample fractionation or a specific feature of the SAX method. The same mixture of α-/β-casein and myoglobin peptides, containing pSer, pThr and pHis peptides as well as non-phosphorylated peptides was fractionated by high pH RP chromatography. As for the SAX fractions, the high pH RP fractions were pooled and analysed by LC-MS/MS, followed by label free quantification with PEAKS. The ion current for all peptides in each fraction was summed and the percentage attributable to either phosphorylated or non-phosphorylated peptides was calculated (Figure 4.4 a). The total signal for a given peptide was summed across all fractions, and the percentage observed in each fraction plotted to compare fractions in which pHis peptides and their non-phosphorylated equivalents elute. Results for two representative peptides are shown (Figure 4.4 b and c).

Unlike for SAX fractionation where the later fractions are enriched for phosphopeptides, high pH fractionation results in elution of phosphorylated peptides predominantly in the first three fractions. Elution of phosphopeptides at the start of the gradient is unsurprising given that the addition of a phosphate group makes a peptide more hydrophilic and therefore reduces its affinity to a C18 column. The percentage signal from phosphopeptides in these early fractions does not exceed 80%, whereas with SAX fractionation the later fractions exhibit up to 100% phosphopeptide signal. These data demonstrate that the phosphopeptide enriched fractions observed following SAX fractionation are not simply a result of fractionation, i.e. spreading out the peptides across multiple fractions, rather it is the charge-based separation inherent to SAX that results in the desired phosphopeptide enrichment effect.

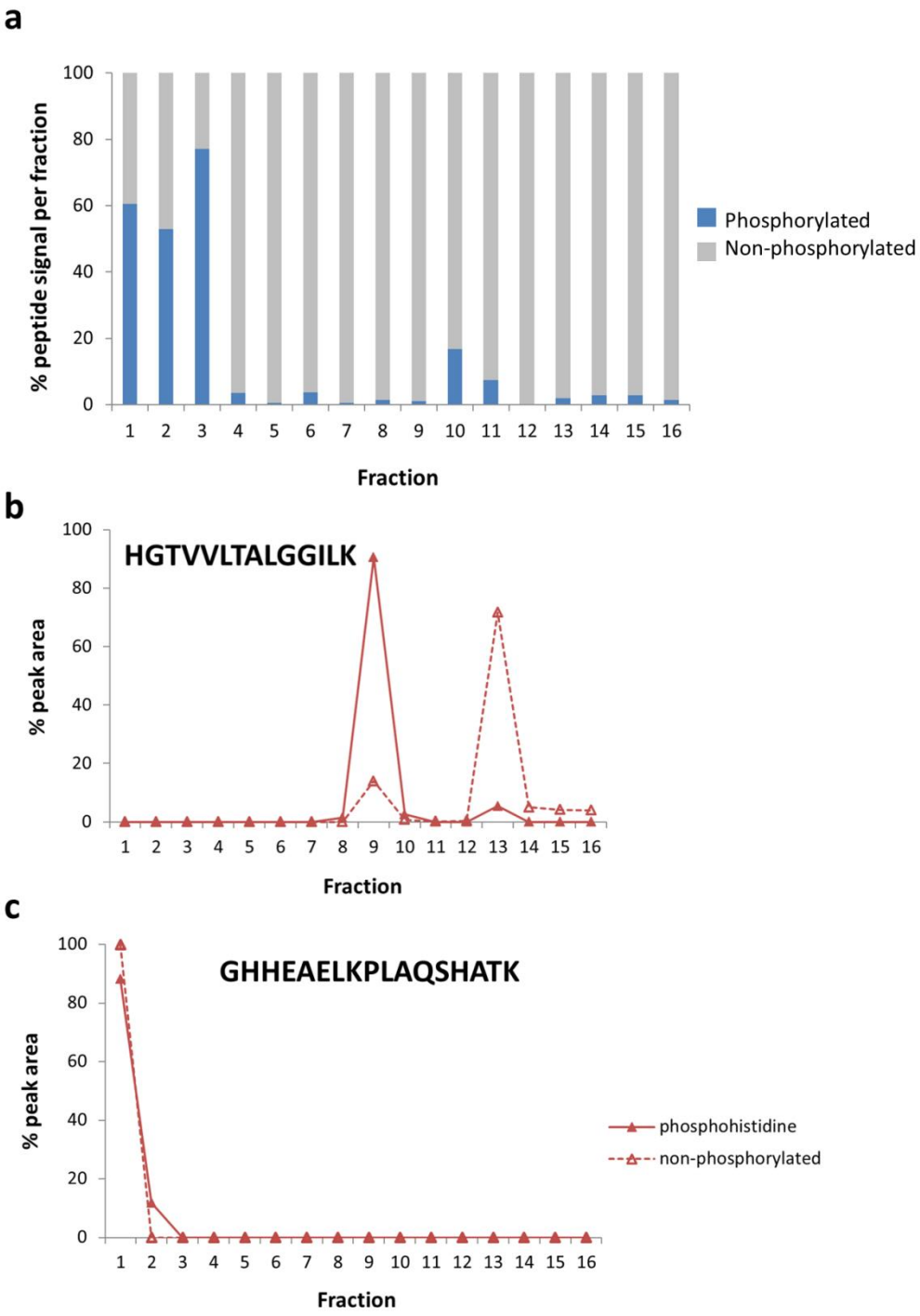


Figure 4.4. High pH reversed-phase fractionation of tryptic peptides of α-/β-casein and myoglobin a) Percentage of the total signal intensity attributed to phosphopeptides (blue bars) and non-phosphorylated peptides (red line) in each fraction following LCMS/MS analysis and label free quantification using PEAKS, with phosphopeptide signal predominantly in first three fractions. **b)** and **c)** Label free quantification of two representative pHis peptides (solid lines) and their non-phosphorylated equivalents (dashed lines) in individual fractions, expressed as a percentage of the total ion current across all fractions.

Inspection of label free quantification data for individual pHis peptides and their non-phosphorylated equivalents following high pH RP fractionation, highlights two main problems. Two out of the five pHis peptides of myoglobin eluted after the first three fractions, so are in fractions that predominantly consist of non-phosphorylated peptide signal. In the example shown the HGTVVLTALGGILK peptide (Figure 4.4 b) elutes in fraction 9, which contains 99% signal intensity from non-phosphorylated peptides (Figure 4.4 a). Elution of pHis peptides in fractions containing predominantly non-phosphorylated peptides would become more problematic as sample complexity increases, as phosphopeptides in these fractions could result from the same ion suppression effects commonly observed for non-enriched samples, thus hindering the identification of phosphopeptides in these fractions. For two of the five pHis peptides of myoglobin it was additionally observed that both the phosphorylated and non-phosphorylated version eluted in the same fraction, as demonstrated by the example peptide GHHEAELKPLAQSHATK (Figure 4.4 c), indicating that this method was not able to separate the two forms. Again, poor separation of phosphorylated and non-phosphorylated peptides would undoubtedly hinder analysis of more complex samples.

Despite high pH reversed-phase chromatography being proven to be effective for pre-fractionation of samples, including prior to (or after) TiO_2 enrichment for increased coverage of the phosphoproteome[174], it is not suitable as a stand-alone strategy for fractionation where the goal is to obtain phosphopeptide enriched fractions. For the purpose of phosphopeptide enrichment, the SAX method developed in this work is far superior.

4.2.3 Peptide recovery by SAX fractionation

An assessment was made regarding the recovery of pHis peptides throughout the sample preparation pipeline thus far developed, i.e. SAX fractionation, pooling and drying of fractions, and rapid C18 desalting (<10 minutes) prior to LC-MS/MS analysis. Recovery of (phospho)peptides was primarily assessed in order to determine the potential loss through dephosphorylation that could be occurring for pHis containing peptides, analogous to the assessment made of the efficiency of other phosphopeptide enrichment strategies (*Chapter 3. Results I*). It was also deemed important to establish the overall losses that may be incurred as a result of a procedure with multiple sample handling steps. Using the label free quantification data generated with PEAKS, the total signal across all fractions for a given peptide was compared to the signal for the same peptide in an unfractionated sample; the amount of unfractionated sample analysed by LC-MS/MS was equivalent to the total amount analysed across all 16 fractions. Peptide recovery is shown for a representative selection of pHis, pSer/pThr and non-phosphorylated peptides from α-/β-casein and myoglobin (Table 4.2).

Table 4.2 Recovery of tryptic peptides of myoglobin and α-/β-casein following SAX fractionation and C18 desalting. Recovery calculated by comparison of sum of peptide intensity across all fractions compared to signal intensity of peptide in unfractionated sample, following LC-MS/MS analysis and label free quantification using PEAKS. Phosphorylated residues are underlined.

	Peptide Sequence	**m/z**	**% recovery**
Myo (pHis)	YLEFISDAIIpHVLHSK	655.6	16.4
	VEADIAGpHGQEVLIR	843.9	31.6
	LFTGpHPETLEK	676.3	35.1
Myo (non-phosphorylated)	YLEFISDAIIHVLHSK	472.0	17.9
	VEADIAGHGQEVLIR	536.3	40.3
	LFTGHPETLEK	424.6	19.2
Myo (no histidine)	YKELGFQG	471.2	36.5
Casein (pSer/Thr)	VPQLEIVPNpSAEER	830.9	50.5
	TVDMESpTEVFTK	733.8	43.9
	FQpSEEQQQTEDELQDK	687.9	21.4
Casein (non-phosphorylated)	VPQLEIVPNSAEER	790.9	75.4
	TVDMESTEVFTK	693.8	22.2
	FQSEEQQQTEDELQDK	991.4	101.1

Recovery is comparable between all peptides regardless of whether they are phosphorylated or not, or the residue of phosphorylation. Only one of the peptides analysed exhibited 100% recovery, with an average recovery of 39% across the 13 representative peptides. Less than 100% recovery is perhaps to be expected given the number of sample preparation steps involved, each with the potential for samples losses to occur. However, encouragingly the recovery of pHis peptides is similar to that of other pSer/pThr phosphopeptides indicating that extensive dephosphorylation will not significantly hamper this analysis, as was the case for the previously trialled enrichment techniques where recovery of pHis peptides was often 0% (*Chapter 3. Results I*). The observed sample losses can be tolerated for the purposes of this work, although further consideration would be required if utilising this strategy for a quantitative rather than qualitative analysis.

4.2.4 SAX fractionation of cell lysate

Following on from the successful application of the SAX method for the analysis of model pHis peptides from myoglobin, the next step was to ensure compatibility of the method with analysis of more complex samples such as a whole cell lysate. Peptides of α-/β-casein and myoglobin were analysed by SAX at pH 6.8 following digestion in urea (8 M original concentration, diluted to 2 M prior to addition of trypsin), either as the simple peptide mixture alone or spiked into U2OS cell lysate. Each fraction was analysed by LC-MS/MS and label-free quantification performed as previously described using PEAKS. For any given peptide the total signal intensity across all fractions was summed, and the amount in each individual fraction then displayed as a percentage of this total amount. The elution of pHis peptides and their non-phosphorylated equivalents are compared across the various conditions (Figure 4.5).

In the all cases the peptides elute in the same (or an adjacent) fraction regardless of the addition of the urea or the presence of peptides from a whole cell lysate. For the sample digested with urea and a whole cell lysate all pHis peptides are observed in different fractions to their non-phosphorylated equivalents. The effective separation of pHis and non-phosphorylated peptides indicates that the method is robust, and the sample preparation techniques proposed for use with whole cell lysate will be suitable for fractionation of these samples using the SAX method.

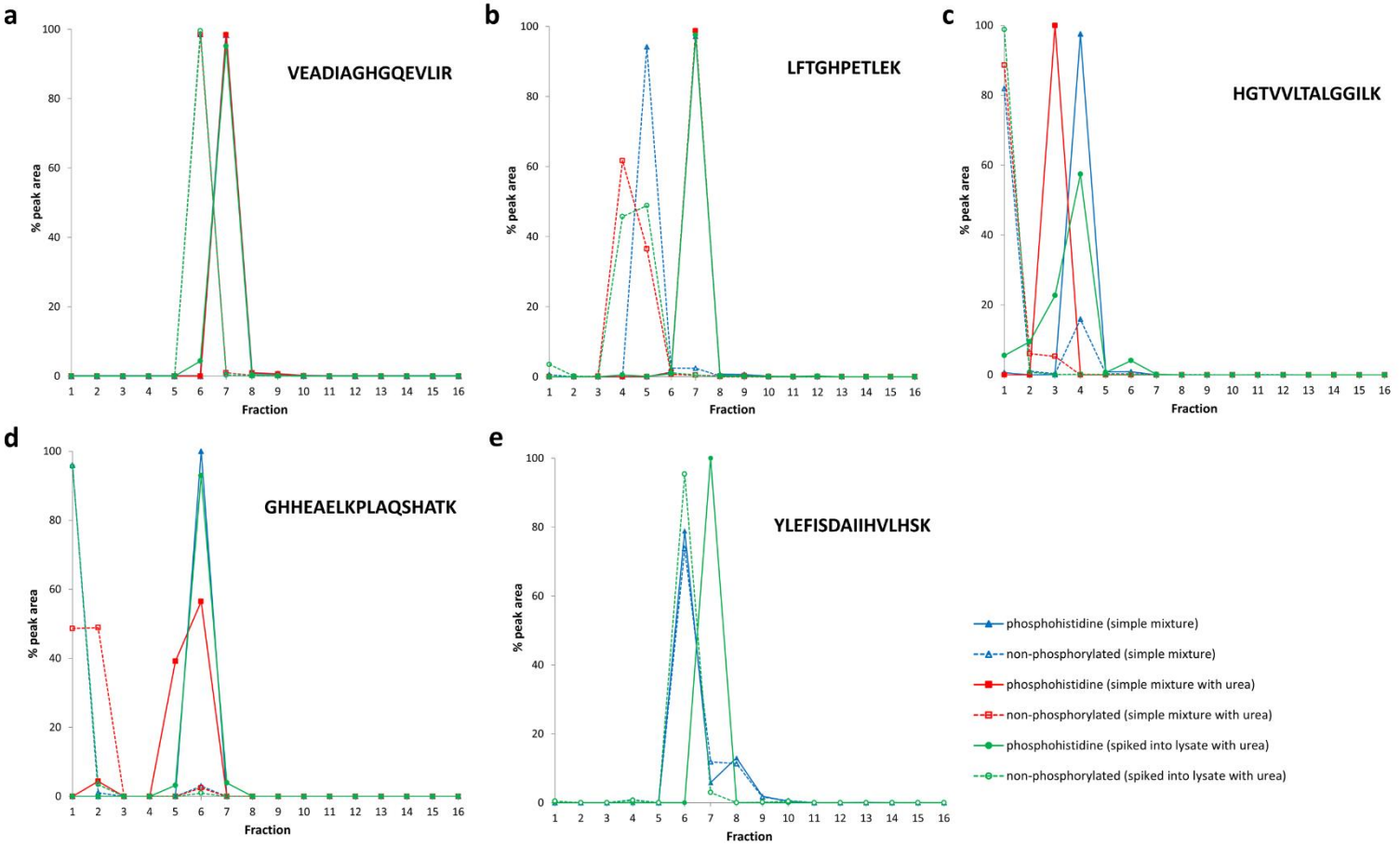


Figure 4.5 Comparison of pHis peptide separation by SAX at pH 6.8 following digestion under different conditions. Label free quantification of phosphohistidine containing peptides (solid lines) and their non-phosphorylated equivalents (dashed lines) in individual SAX fractions, expressed as a percentage of the total ion current across all fractions, following digestion in AmBic (blue triangle) digestion in urea (red square) and digestion in urea with U2OS lysate (green circle). YLEFISDAIIHVLHSK peptide (e) not detected in sample digested in urea.

Fractionation of a U2OS cell lysate (1 mg) was performed using the SAX method previously described. Sample loading for LC-MS/MS was optimised by increasing the amount of the later fractions analysed to achieve approximately equal base peak chromatogram intensities for all fractions. LC-MS/MS data was generated on the Orbitrap Fusion mass spectrometer using a HCD with neutral loss triggered EThcD method. Data was searched using the MASCOT search engine with oxidation (M) and phosphorylation (ST, Y and H) as variable modifications. Peptides are reported at 1% FDR. An in-house perl script was used to determine unique peptide IDs across all fractions, where the highest scoring (MASCOT ion score) peptide was kept. The number of phosphorylated and non-phosphorylated peptides in each fraction was plotted to assess the extent of phosphopeptide enrichment across the SAX gradient (Figure 4.6).

In the later SAX fractions the percentage of phosphopeptides approaches 45%. Indicating that these fractions are considerably enriched for phosphopeptides compared to a non-fractionated sample; phosphopeptide IDs are typically <1% of total peptide IDs in non-fractionated U2OS cell lysate digest analysed in the same way. An attempt was made to improve coverage of phosphopeptides by removing the fraction pooling step and instead analysing the 46 SAX fractions by LC-MS/MS individually. The average percentage of phosphopeptides in the later fractions (39-46, approximately corresponding to fractions 15 and 16 of the pooled analysis) was 50% representing only a very modest improvement. A 37% increase in phosphopeptide IDs was obtained by analysis of individual rather than pooled fractions, again representing a modest improvement. However, to achieve these improvements requires a vast increase in sample preparation and MS instrument time; LC-MS/MS analysis of 46 samples with a 2 hour method approaches 4 days, which would soon become unfeasible for experiments with multiple biological conditions and replicates.

For the data obtained by analysis of 16 pooled fractions, the number of unique phosphopeptides in each fraction was further separated into the different phosphosites, with localisation determined by MASCOT (Figure 4.7). Analysis of the 16 SAX fractions revealed 29 endogenous pHis containing peptides, 12 of which were phosphorylated only on histidine. A total of 2121 phosphopeptides were identified across all the fractions, the majority of which were phosphorylated on serine (1911, 90%). The SAX fractionation strategy was thus shown to be suitable for identification of pHis peptides alongside other sites of phosphorylation, representing a theoretically unbiased approach to phosphopeptide analysis.

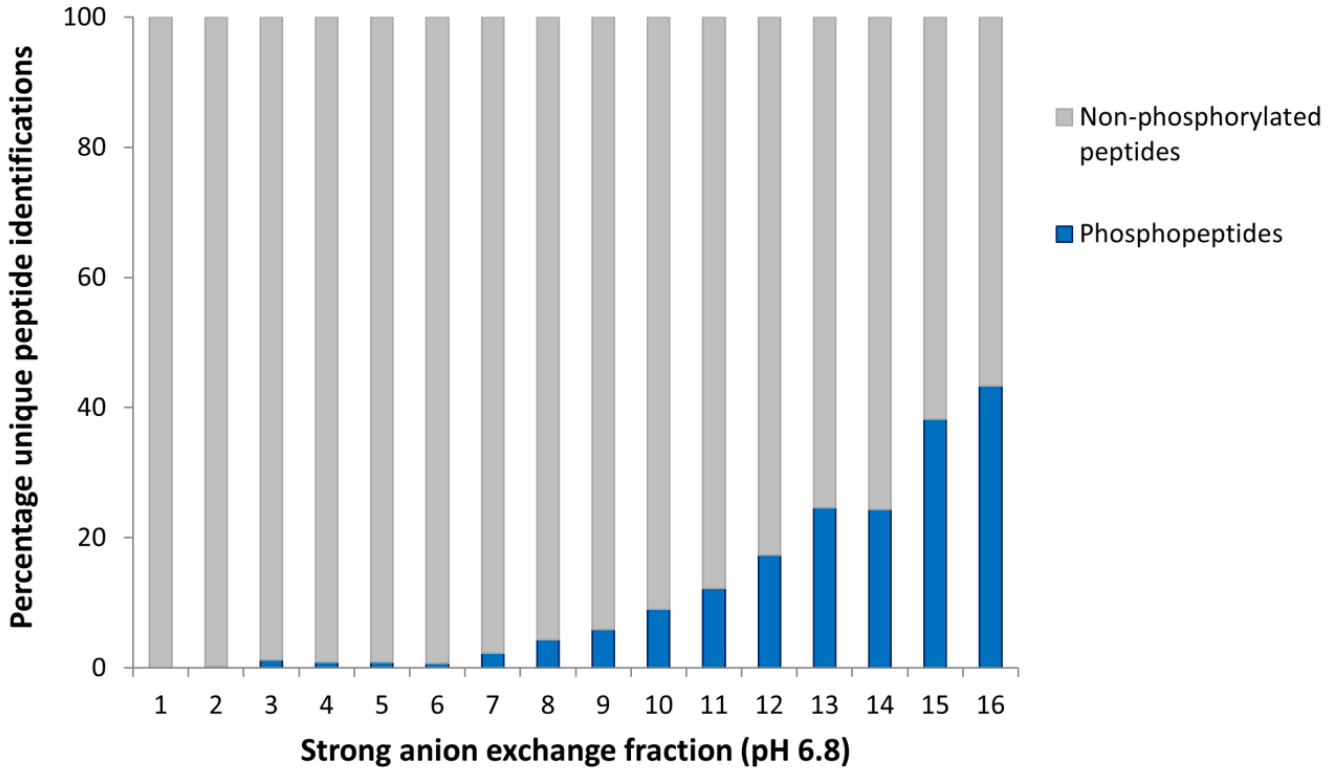


Figure 4.6 Percentage of unique phosphorylated and non-phosphorylated peptides identified in each fraction following SAX fractionation of U2OS lysate at pH 6.8. Peptides from tryptic digest of U2OS lysate (1 mg) separated by SAX at pH 6.8, with 16 pooled fractions analysed by LC-MS/MS using the Orbitrap Fusion, then searched with MASCOT at 1% FDR. Later fractions show a considerable degree of enrichment, containing up to 45% phosphopeptides.

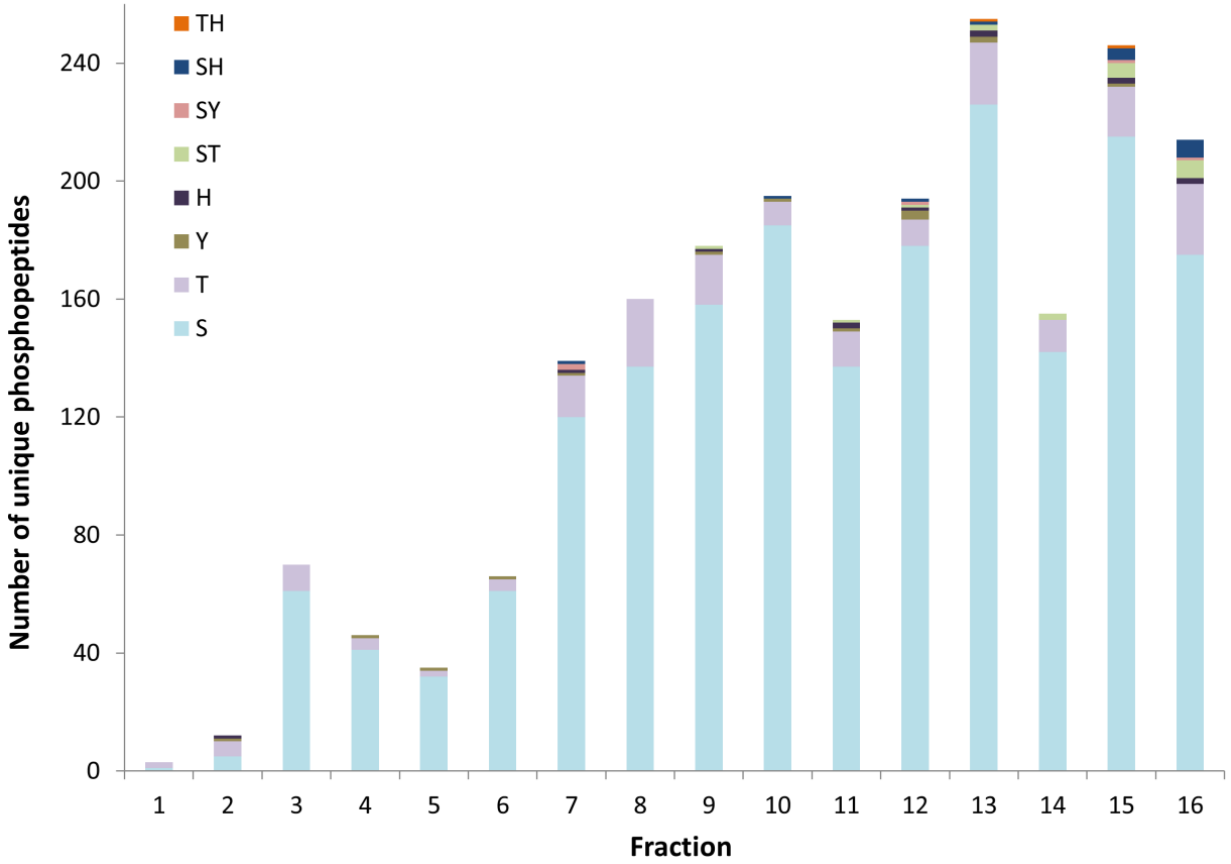


Figure 4.7 Number of unique phosphopeptides identified in each fraction following SAX fractionation of U2OS lysate at pH 6.8. Later fractions show increased numbers of phosphopeptides compared to early fractions. Site of phosphorylation determined by MASCOT, with the majority of sites identified as pSer and a total of 29 endogenous pHis peptides identified.

4.2.5 SAX fractionation using StageTips

The SAX phosphopeptide separation method as it has been developed thus far uses 1 mg of digested cell lysate, which may become problematic if only limited sample amounts are available, for example from experiments involving biological tissue or primary cell lines. Utilising a StageTip format rather than a large chromatography column would enable smaller sample amounts to undergo SAX fractionation, as well as significantly reducing the time taken to fractionate each sample. A rapid SAX fractionation strategy would be especially beneficial for studies with many samples, as currently it requires at least a day to perform the fractionation and subsequent drying and desalting of the 16 fractions for a single sample.

StageTips are routinely used as a simple method of sample clean-up and fractionation[248, 249]. For this experiment the tips were packed with two discs of anion exchange material, and a step-elution fractionation performed. Based on preliminary experiments which revealed that all peptides (phosphorylated and non-phosphorylated) were eluting in early (low salt concentration) fractions, the concentration of triethylammonium phosphate was reduced such that the concentration in the final fraction was ~200 mM compared to 300 mM in the column method. The method was tested with 40 µg of a simple mixture of pHis-myoglobin and α-/β-casein peptides, a much lower amount than is required for SAX fractionation by the column method. Each fraction was desalted and analysed by LC-MS/MS, using the Bruker AmaZon mass spectrometer. The low-resolution data generated by the Bruker Amazon mass spectrometer is not compatible with label-free quantification using PEAKS therefore assessment of fractionation efficiency is based on peptide numbers alone. The percentage of unique phosphorylated and non-phosphorylated peptides in each fraction was calculated, with the average across two technical replicates plotted in Figure 4.8.

A moderate degree of phosphopeptide enrichment is observed with the later fractions containing ~40% phosphopeptides. A direct comparison to the SAX column data is not possible due to different methods of LC-MS/MS analysis, but a greater degree of phosphopeptide enrichment using the StageTip method would be desirable, especially given that the degree of enrichment would be expected to drop further for a complex sample.

Inferior performance of the StageTip method compared to the SAX column based method may be expected as the larger packed column provides considerably better resolution than the StageTip format, meaning less interference in later fractions from non-phosphorylated peptides. It was observed in the StageTip data that the majority of peptides elute in at least two or more fractions, with a few being identified in up to six fractions. Perhaps performing multiple elution steps at the same salt concentration would reduce the likelihood of peptides spreading across multiple fractions. Additionally, using anion exchange resin packed into a small column, either home-made or a commercial product, rather than the discs of anion exchange material may provide improved performance on a smaller scale. Techniques to achieve comparable results to the SAX method developed using the large-scale column but on a smaller scale still require further investigation.

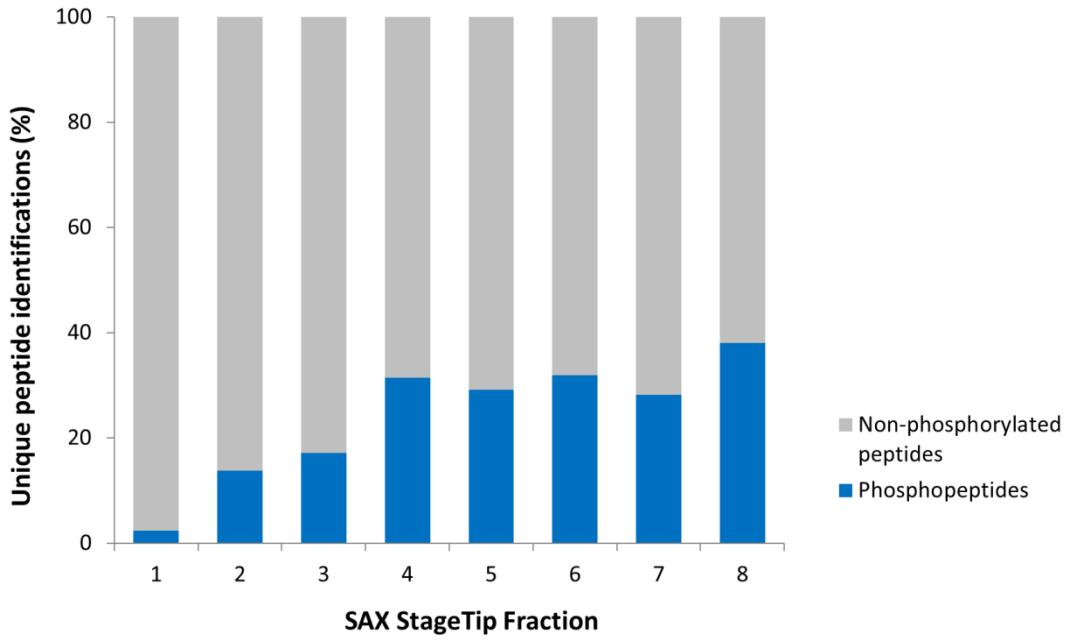


Figure 4.8. Percentage of unique phosphorylated and non-phosphorylated peptides identified in each fraction following SAX StageTip fractionation of α-/β-casein and myoglobin peptides at pH 6.8. Eight fractions were collected by step-elution with increasing salt concentration, desalted and analysed by LC-MS/MS using the Bruker AmaZon. Data searched with MASCOT at 1% FDR and unique peptides determined in each fraction. Later fractions (6 to 8) contain 30-40% phosphopeptides, compared to 2% for fraction 1.

4.3 Conclusions

A SAX fractionation strategy for phosphopeptide enrichment based on salt gradient elution has been developed. Assessment of the elution of pHis peptides and their non-phosphorylated equivalents with buffers at three different pH values (pH 6. 6.8 and 8) revealed pH 6.8 to be optimal for the analysis of pHis-containing peptides. For a simple mixture SAX fractions at the end of the gradient contain up to 100% phosphopeptide signal. The recovery of pHis peptides by this method is up to 35% which is comparable to recovery of other peptides (pSer/pThr and non-phosphorylated) using this workflow and much improved compared to other enrichment strategies previously assessed in *Chapter 3. Results I*. Unlike other enrichment strategies whereby the "unwanted" material is discarded and not analysed this SAX fractionation method allows for LC-MS/MS analysis of the whole sample. With this SAX method all sites of phosphorylation can be assessed, representing a novel unbiased approach to phosphopeptide analysis. The method has subsequently been termed UPAX (Unbiased Phosphopeptide enrichment by strong Anion Exchange).

The ability of the UPAX method to identify pHis peptides in a complex cell lysate has been clearly demonstrated, with later fractions containing up to 45% phosphopeptides and 29 endogenous pHis peptides identified following SAX fractionation of a U2OS cell lysate. The method will therefore be utilised in further studies to examine the extent of acid-labile phosphorylation in mammalian cells, and the types of proteins on which these modifications occur. The fact that this method may not be applicable for studies where sample is limited has been addressed by modification of the technique for SAX fractionation in StageTips, using discs of anion exchange material packed into a pipette tip and a step-elution of increasing salt concentration. Although some degree of enrichment is observed in the later fractions, the method still requires further development in order to achieve a comparable performance to the large packed column with which the method was originally developed.

The data analysis strategy used thus far has been relatively simple, using only MASCOT for phosphopeptide localisation, with no consideration given for confidence of the sites of modification, for example by application of a phosphosite localisation scoring strategy[209]. It will be important to develop a data analysis pipeline to complement the UPAX strategy, such that confident pHis peptide identifications can be reported. Application of well-

thought-out criteria for pHis site localisation will be crucial as further studies to characterise pHis sites and their possible functions are undertaken.

Chapter 5. Results III: Phosphohistidine in human cells

5.1 Introduction

This chapter describes the implementation of the successfully developed UPAX strategy for the separation and characterisation of pHis peptides, derived from HeLa cell lysates, at non-acidic pH. Whilst there is precedent for using pervanadate to inhibit Tyr phosphatases for global pTyr site mapping studies[250, 251], no such inhibitors have yet been characterised for His phosphatases. Instead, siRNA knockdown of PHPT1, one of two known His phosphatases, was conducted in HeLa cells with the intention of increasing the basal level of His phosphorylation.

Phosphorylation of His was first reported in mammalian cells over 50 years ago[10], and since then there has been growing evidence for the roles of pHis in mammalian signalling. Histone H4 kinases activity appears to be associated with cell growth and development[41-43], and NME1 and NME2, the only reported His kinases to date, are implicated in cancer and tumour metastasis[25-27]. His phosphorylation has also been implicated in differentiation of PC12 neuronal cells[23], T-cell signalling[30, 34] and glycoprotein VI signalling in platelets and bone remodelling[21, 22]. However, widespread characterisation of His phosphorylated proteins and their functions in mammalian systems is still lacking. The work presented in this chapter thus aimed to address this issue by application of the UPAX strategy, enabling characterisation of hundreds of sites of His phosphorylation in human cells.

As described in *Chapter 1. Introduction*, phosphosite localisation presents a significant challenge in phosphoproteomics experiments. Therefore, it was vital to implement a robust phosphosite localisation strategy into the pHis analysis pipeline, such that pHis sites could be reported with confidence. From the many available approaches for site localisation[209], either integrated into search engines or available as post-search engine tools, ptmRS was chosen for this study. However, an additional strategy to confirm sites of His phosphorylation was also considered. Oslund *et al.*[252] studied gas-phase fragmentation of pHis peptides and thus proposed that a characteristic neutral loss pattern was indicative of His as the site of phosphorylation. LC-MS/MS data was interrogated to determine whether the triplet neutral loss strategy could be utilised for high-throughput characterisation of

pHis peptides, by virtue of improving confidence in localisation of His as the site of phosphorylation, as opposed to another amino acid.

The results presented in this chapter represent the largest number of localised pHis sites to be reported to date, which provided a unique opportunity to characterise motifs for His phosphorylation. Additionally, the pHis proteins identified in these samples were assessed in terms of functional annotation and enrichment analysis. Widespread His phosphorylation has been revealed in human cells, and novel biological roles for pHis were thus considered.

5.2 Results and Discussion

5.2.1 siRNA knockdown of PHPT1

PHPT1 knockdown was performed in HeLa cells with the aim of increasing levels of pHis. PHPT1 is ubiquitously expressed in mammalian cells[253], and the HeLa cell line has been shown to contain pHis-containing proteins by western blotting and/or immunofluorescence staining with 1- and 3-pHis antibodies[227]. Additionally, PHPT1 has been shown to be elevated in other cancer cell lines; increased PHPT1 expression correlates with increased tumour size in renal cell carcinoma and increased cell migration and invasion in lung cancer[254, 255].

Knockdown of PHPT1 with siRNA was performed in triplicate, alongside a positive control Lamin A/C knockdown (Figure 5.1). All bioreplicates demonstrate successful knockdown of PHTP1 compared to control lysates treated with non-targeting (NT) siRNA. The two conditions are hereby referred to as 'PHPT1 siRNA' and 'NT siRNA'.

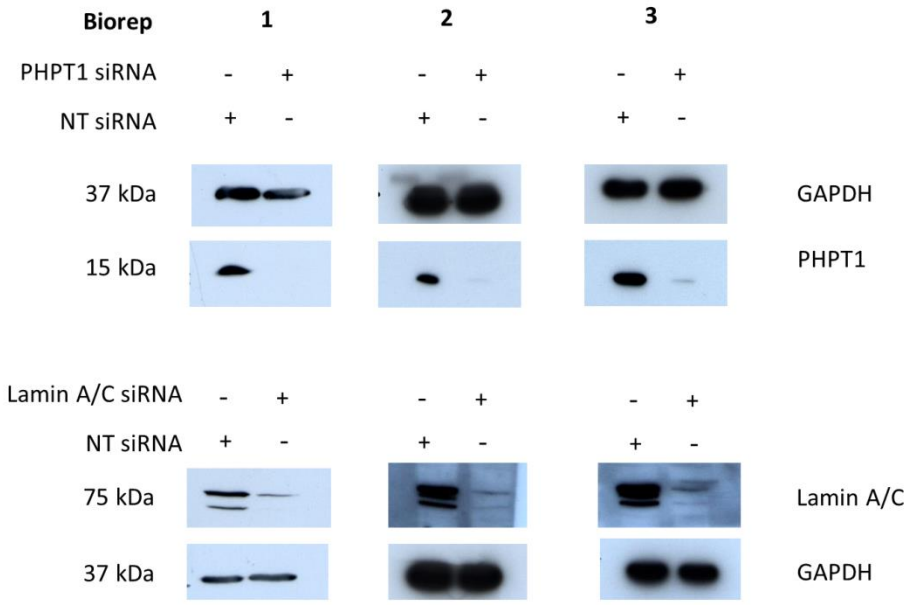


Figure 5.1 Knockdown of PHPT1 in HeLa cells. In all three bioreplicates almost complete reduction in levels of PHPT1 (15 kDa) in cells treated with PHPT1 siRNA for 48 hours is observed, with GAPDH (37 kDa) loading control also shown. Levels of LaminA/C (75 kDa), which was used as knockdown control, are also reduced after 48-hour treatment.

5.2.2 Implementation of SAX fractionation strategy

The UPAX strategy successfully developed for separation and characterisation of phosphopeptides in complex cell lysates (*Chapter 4. Results II*), was applied to the HeLa cell lysates collected following the siRNA knockdown experiment previously described. The general workflow implemented for this analysis is outlined in Figure 5.2.

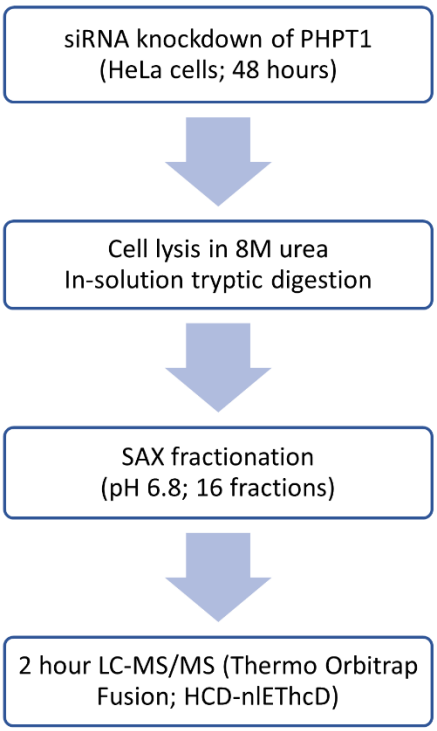


Figure 5.2. Workflow for PHPT1 knockdown and SAX experiment. Following treatment with PHPT1 or non-targeting siRNA for 48 hours cells were lysed and proteins digested with trypsin. Peptides were fractioned by SAX at pH 6.8 with salt gradient elution and the resulting 16 fractions were analysed by LC-MS/MS using the Thermo Orbitrap Fusion.

Cells were lysed and digested with trypsin, under non-acidic conditions and without exposure to elevated temperature. Peptides were fractionated by SAX at pH 6.8, and 16 fractions analysed by LC-MS/MS, using a Thermo Orbitrap Fusion mass spectrometer. Fragmentation was performed by HCD with neutral loss triggered EThcD. According to this fragmentation scheme a neutral loss of 98 Da from the precursor in the HCD fragment ion spectrum, triggers the precursor to undergo a second round of fragmentation by ETD with supplemental HCD activation to produce b/y and c/z ions. HCD-neutral loss triggered EThcD

provides a good balance between rapid analysis achieved using HCD and superior phosphosite localisation afforded by an ETD-based fragmentation method[74, 200]. Representative base peak chromatograms (Figure 5.3 a) indicate different profiles observed for LC-MS analysis of different fractions throughout the SAX gradient. Loading of each fraction onto the LC-MS/MS system was adjusted following estimation of peptide amount by Nanodrop and test injections, such that the base peak intensity of each fraction was approximately equivalent.

Proteome Discoverer (PD) was used to implement a workflow for database searching and phosphosite localisation (Figure 5.3 b). Raw files were first processed using MSConvert to perform MS2 level de-isotoping, and the resulting mzML file then provided as the input for the PD workflow. Each file was split into spectra generated by HCD or EThcD, with each type processed separately due to the different settings required to search each type of data. For site localisation the ptmRS node was used. This is the most recent version of the PhosphoRS software developed by Taus *et al*[206]. The algorithm assesses all possible sites that can be phosphorylated and assigns a confidence score (between zero and 100) to each one, where the higher the score for a given site the more likely it is to be phosphorylated. Scores >99 indicate very high confidence in the phosphosite localisation, whilst sites with scores >75 are considered to be confidently localised, with this cut-off routinely used for phosphoproteomics studies[94, 256].

Protein and peptide level results were exported from PD for further interrogation using manual methods and/or in-house scripts. A 5% FDR at the peptide level was applied for all exported data. For proteins an additional criteria of ≥2 peptides per protein was applied. From peptide data only rank 1 peptides were considered for further analysis. For phosphopeptide identifications a 1% FLR was applied with ptmRS scores defined as 99.2 for HCD and 90.2 for EThcD data, according to the benchmarking study performed by Ferries *et al*[200].

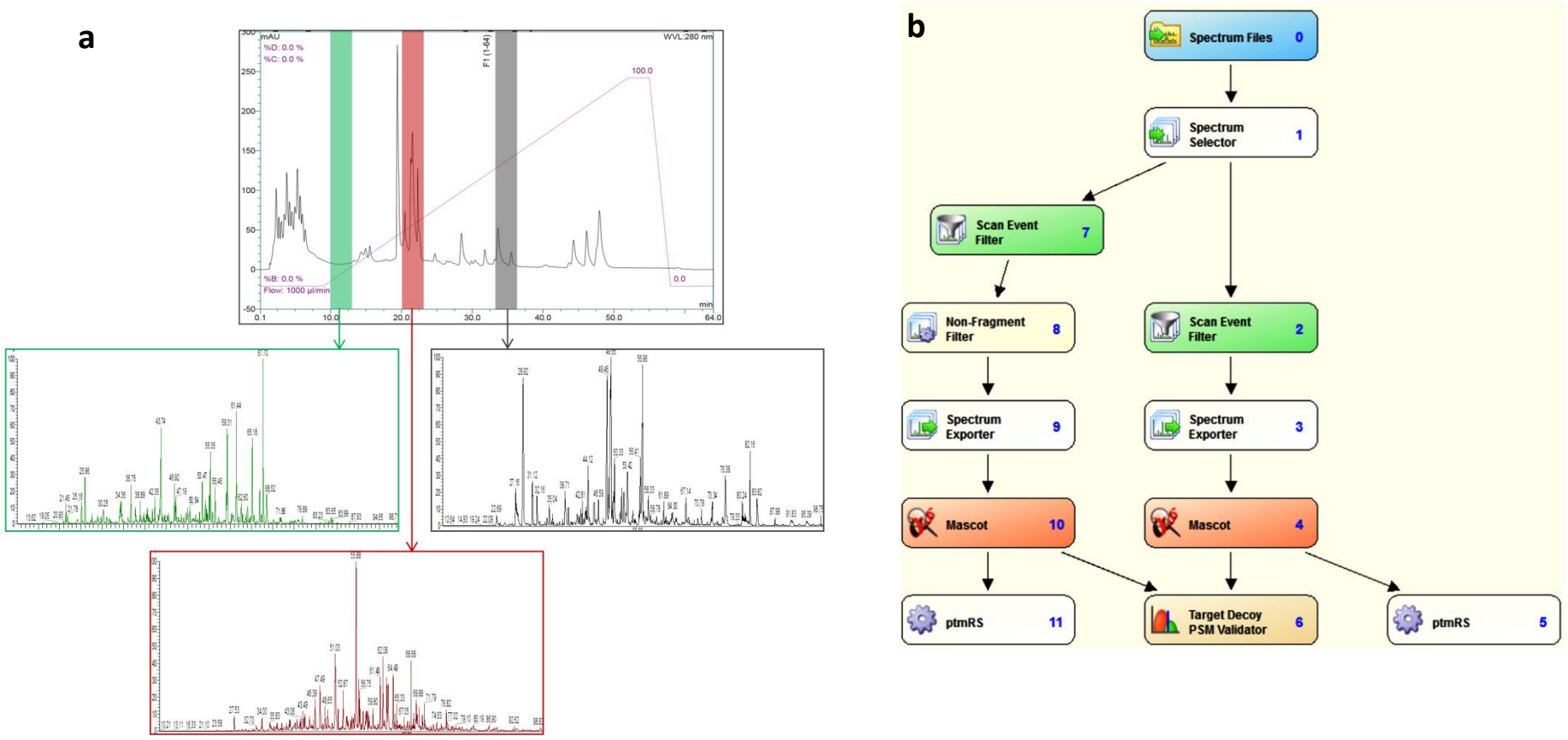


Figure 5.3 LC-MS/MS analysis of SAX fractions and data processing with Proteome Discoverer a) Base peak chromatograms shown for select SAX fractions (3- green; 6- red; 10- grey) following LC-MS/MS using an Orbitrap Fusion mass spectrometer; peptides fragmented by HCD with EThcD triggered by neutral loss of 98 Da from the precursor ion. **b)** Proteome Discoverer data processing pipeline; tandem mass spectra are separated according to fragmentation type prior to searching with MASCOT and application of ptmRS for phosphosite localisation.

An initial assessment was made of the number of phosphorylated and non-phosphorylated peptides in each fraction to establish the extent of phosphopeptide enrichment achieved with the SAX fractionation strategy. Unique peptides and phosphopeptides were determined across all fractions, with the highest scoring (MASCOT ion score) peptide kept in the data set. The number of phosphorylated and non-phosphorylated peptides in each fraction is displayed as a percentage of total number of peptides identified. Data for one representative bioreplicate is shown in Figure 5.4, with graphs for all bioreplicates shown in the Appendix. As previously observed for fractionation of a U2OS lysate (*Chapter 4. Results II*) the later fractions contain ~50% phosphorylated peptides, indicating a significant degree of enrichment compared to an unfractionated sample. All bioreplicates showed a similar trend for fractionation, with an average of 47% phosphorylated peptides in fraction 16 across the six sets of fractions, thus demonstrating that the SAX fractionation has performed as expected for these samples.

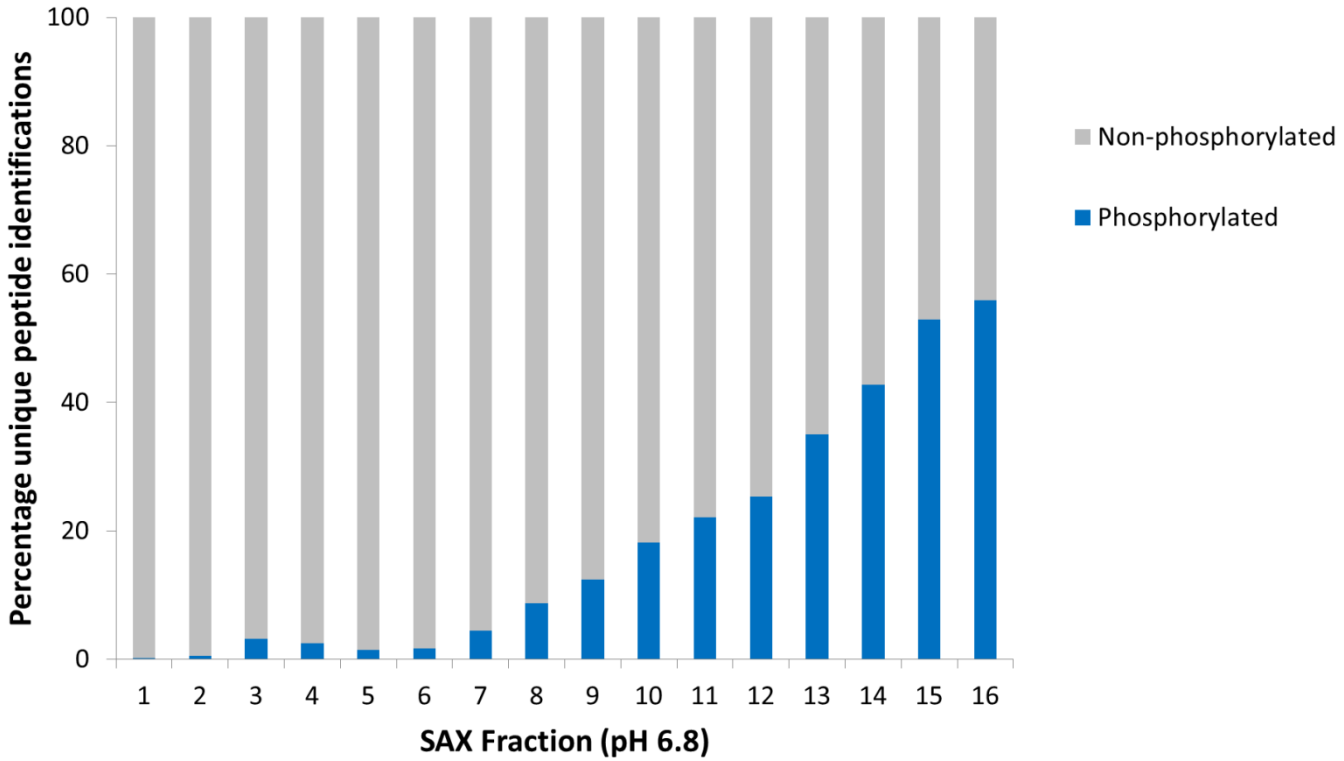


Figure 5.4 Percentage of unique phosphorylated and non-phosphorylated peptides identified in each fraction following SAX fractionation of HeLa lysate at pH 6.8. HeLa cells treated with non-targeting siRNA and analysed according to previously described workflow (Figure 5.2). Proteome Discoverer pipeline used to search data with MASCOT, with peptide IDs exported at 5% FDR and unique peptides determined using an in-house perl script. Later fractions show a considerable degree of enrichment, containing approximately 50% phosphopeptides.

5.2.3 Global effects of PHPT1 knockdown

Protein identifications across the three bioreplicates for each condition show a high degree of overlap (Figure 5.5 a); for PHPT1 siRNA samples a maximum of 10% of the identified proteins were unique to any given individual bioreplicate (10.5%, 7.4%, 7.3%), and 62% of the 5232 proteins identified were common between the three replicates. Similarly, for the NT siRNA control samples, 63% of the 5353 proteins identified were common across the three bioreplicates, with a maximum of 11% (11%, 8%, 7%) being unique to any individual bioreplicate analysis. Overlap between bioreplicates was also considered for non-phosphorylated peptides (Figure 5.5 b): a maximum of 15% (of 90643) or 19% (of 96363) of peptides were unique to a single bioreplicate for the PHPT1 siRNA or NT siRNA treated sample respectively. For phosphopeptides (Figure 5.5 c), the numbers were a slightly more variable, with up to 27% (of 6574) or 26% (of 7111) of identified phosphopeptides being unique to a single bioreplicate for the PHPT1 siRNA or NT siRNA samples respectively.

To assess common identifications between the PHPT1 siRNA and NT siRNA samples, proteins identified in two or more bioreplicates were considered. 83% of the 4580 unique proteins identified across the two conditions, were common, indicating a high degree of overlap. Additionally, 68% of peptides and 55% of phosphorylated peptides were common between the two conditions, considering only those peptides that were identified twice or more across the dataset. Independent of assessment criteria, greater numbers were identified in the NT siRNA compared to the PHPT1 siRNA sample.

There are 340 and 417 protein identification unique to the PHPT1 siRNA or NT siRNA conditions, respectively. Gene ontology (GO) analysis of these proteins was performed using Panther[257] (Figure 5.6). There are no discernible differences in GO terms with respect to biological processes, molecular function or protein class, revealing no significant insights into the global effects of PHPT1 knockdown in HeLa cells.

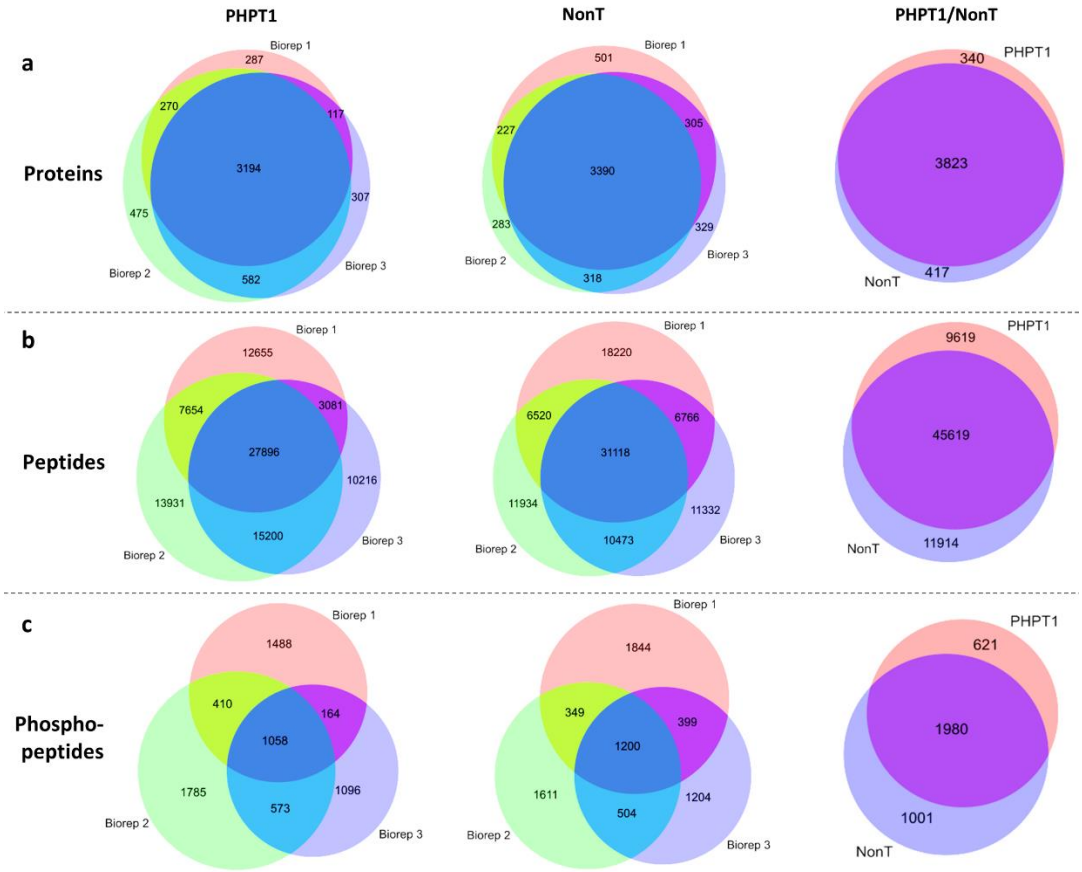


Figure 5.5 Overlap of protein, peptide and phosphopeptide IDs between bioreplicates and conditions a) Protein identifications show high degree of overlap between bioreplicates; 80% of proteins are identified in two or more bioreplicates with 83% overlap between the two treatment conditions **b)** For peptide IDs in two or more bioreplicates there is an overlap of 57% for PHPT1 siRNA samples and 59% for NT siRNA, with 68% overlap between conditions **c)** Overlap of phosphopeptide IDs between conditions is 55%, with 35% of phosphopeptides identified in two or more bioreplicates for both the PHPT1 and NT conditions.

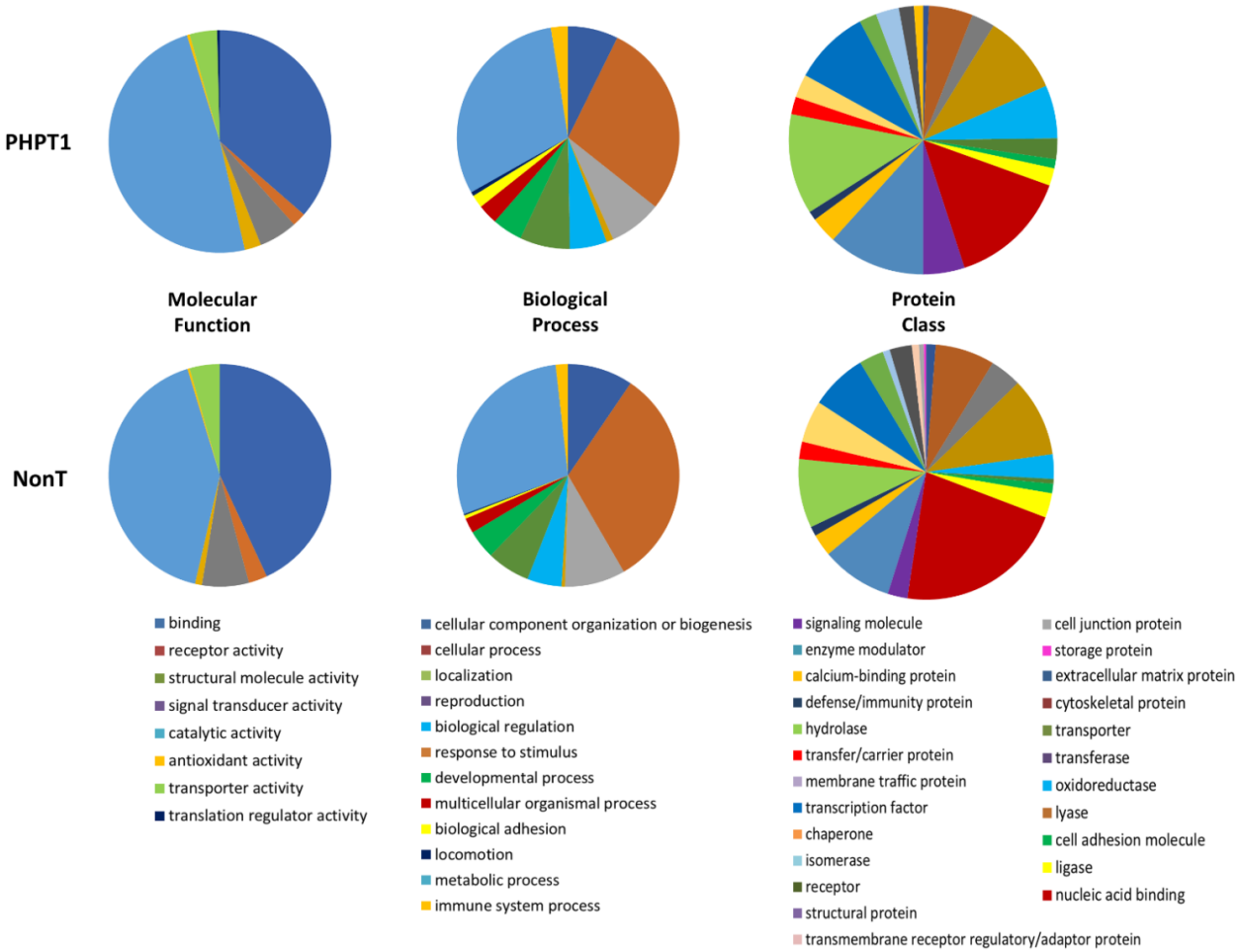


Figure 5.6 GO term analysis of proteins unique to either PHPT1 siRNA or NT siRNA condition. Unique proteins for PHPT1 (340 proteins) and Non-T (417 proteins) show no considerable differences in terms of molecular function, biological process or protein class when searched against the PANTHER classification system.

5.2.4 Evaluation of triplet neutral loss patterns for pHis peptide identification

A crucial aspect of phosphoproteomics analysis is the confident localisation of the phosphate group to a specific residue within a peptide. The use of a tool such as ptmRS can aid in this by providing a confidence score for possible sites of phosphorylation. However, an additional diagnostic strategy to confirm the site of phosphorylation as His, as opposed to another amino acid, would provide an extra level of confidence to these identifications.

It is known that pSer/pThr phosphopeptides can undergo neutral loss of phosphoric acid (Δ98 Da) from the precursor ion during fragmentation[187, 188]. This results in a predominant neutral loss ion in the fragment ion spectrum under CID, and to a lesser extent HCD, fragmentation regimes. In comparison, neutral loss from pTyr peptides is typically observed only at Δ80 Da. The loss of HPO_3 upon fragmentation of the labile P-N bond of pHis phosphopeptides results in a neutral loss of 80 Da, whilst concomitant loss of water from elsewhere in the peptides results in neutral loss of 98 Da. Further loss of water from phosphopeptides undergoing fragmentation gives rise to a Δ116 Da neutral loss which can be observed in pHis, pThr and pSer peptides. It was proposed by Oslund *et al.*[252] that the combination of neutral loss ions observed following fragmentation of a phosphopeptide could be indicative of the amino acid which is phosphorylated, specifically that the 'triplet' neutral loss pattern (Δ80, Δ98 and Δ116 Da) is characteristic of pHis phosphorylation, aiding in the identification of pHis peptides in a phosphoproteomics experiment.

The triplet neutral loss strategy could be invaluable for increasing confidence in the localisation of phosphosites to His residues, particularly in cases where a peptide contains other residues that could be phosphorylated in addition to His. The triplet neutral loss pattern was observed in all fragment ion spectra (generated by HCD) of the pHis peptides of myoglobin (*Chapter 3. Results I, Figure 3.4*), suggesting this was a promising strategy to pursue. To assess the validity of implementing the triplet neutral loss strategy for high-throughput characterisation of pHis-containing peptides, the prevalence of neutral loss ions in the tandem MS spectra of phosphopeptides was assessed using data obtained by fractionation of HeLa lysate and analysis of the 16 fractions by LC-MS/MS using the Thermo Orbitrap Fusion. Data was processed using MASCOT and Proteome Discover as previously described, and tandem MS spectra were subsequently inspected for the presence of product ions corresponding to the expected *m/z* values that would result from neutral losses of 80, 98 and 116 Da from the precursor ion. Each identified peptide was therefore

assigned a 'triplet' score (of 0, 1, 2 or 3) depending on the number of neutral loss ions observed following collisional dissociation.

How does HCD vs CID affect triplet score?

The original paper describing the triplet neutral loss strategy utilised CID fragmentation for the observation of the characteristic neutral loss patterns. Whilst neutral loss ions are a predominant feature of spectra resulting from CID fragmentation, under HCD fragmentation this is not necessarily the case. The higher energy dissociations associated with HCD compared to CID result in a wider range of fragmentation pathways, and thus neutral loss from the precursor is not necessarily dominant product ion[258]. A comparison was therefore made between the triplet scores of phosphopeptides identified in one SAX fraction, following fragmentation by HCD or CID (Figure 5.7). The proportion of phosphopeptides displaying each triplet score is broadly similar for both HCD and CID fragmentation. Crucially, this means that 27% of peptides identified following CID fragmentation do not display neutral loss ions, even when the intensity cut off is set at 2% compared to the base peak ion (as it is for the data shown). Considering each site of phosphorylation separately, CID still does not appear to offer any significant benefits in terms of generation of diagnostic fragment ions compared to HCD. The pattern of triplet scores for pSer/pThr peptides is almost identical for both fragmentation types, and whilst the proportion of pHis peptides displaying the triplet neutral loss is greater when CID fragmentation is used compared to HCD, so is the proportion of pHis peptides with no neutral loss ions at all. Further investigation of the triplet score was based on HCD data only, obtained by LC-MS/MS analysis of 16 SAX fractions.

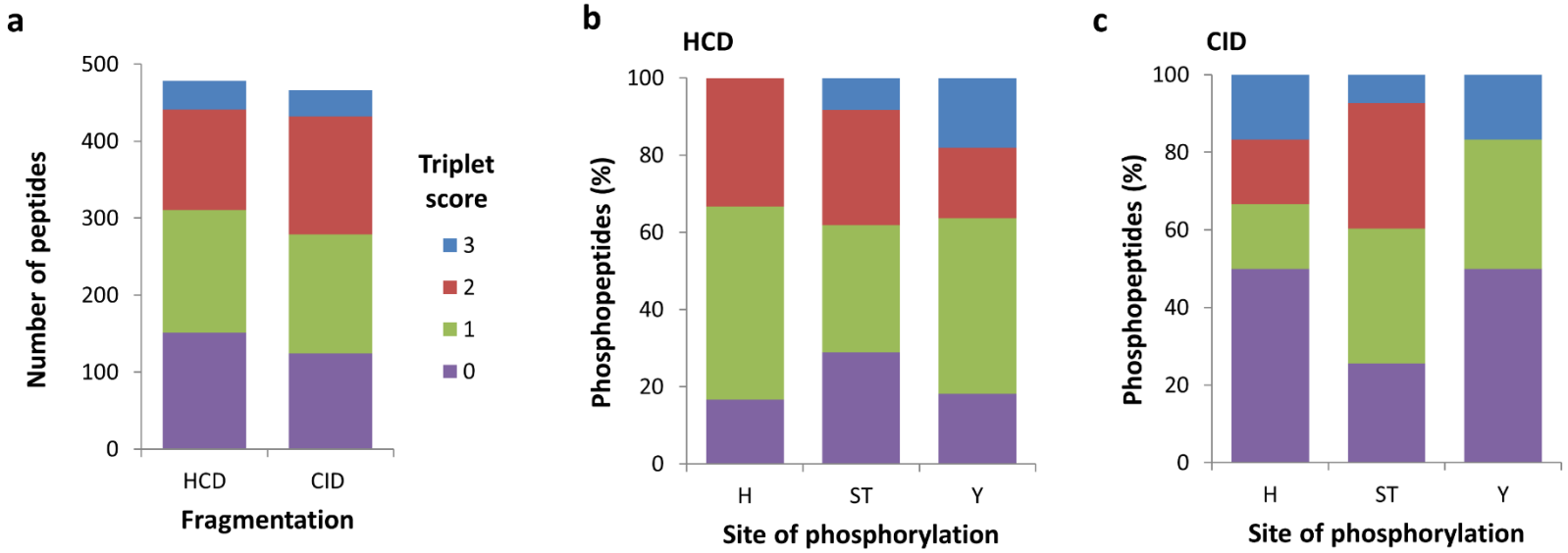


Figure 5.7 Effect of fragmentation method on distribution of triplet scores for phosphopeptides, a) Comparison of triplet scores for all phosphopeptides identified in a single SAX fraction (SAX fraction 13) following fragmentation by either HCD or CID, **b-c)** Distribution of triplet scores for unique singly phosphorylated peptides with ptmRS score ≥75 as a function of site of phosphorylation (pHis, pSer/pThr and pTyr) following **(b)** HCD or **(c)** CID fragmentation.

Is triplet score related to ptmRS score?

Given that ptmRS scores are a measure of confidence in site localisation it was expected that peptides with higher ptmRS scores would display closest adherence to the triplet theory, i.e. the most confidently assigned pHis peptides should exhibit the triplet neutral loss, whereas for low scoring sites there is greater ambiguity over the site of phosphorylation and therefore these peptides may not display the triplet neutral loss pattern, instead displaying a neutral loss pattern indicative of an alternative site of phopshorylation (pSer, pThr or pTyr). The triplet score for all pHis-containing peptides identified across the 16 fractions was thus considered according to four categories of ptmRS score assigned according to different degrees of site localisation confidence (Figure 5.8 a).

Somewhat unexpectedly the number of pHis peptides exhibiting the triplet neutral loss pattern was extremely low, even for phosphopeptides with a ptmRS score greater than 99; only 3% of pHis peptides with ptmRS score >99 have a triplet score of 3. One reason for this could be that the number of phosphosites in the peptide is affecting the neutral loss pattern, i.e. doubly phosphorylated peptides may be more likely to display neutral loss of Δ196 Da (two losses of 98 Da) meaning the expected triplet pattern may not be observed for multiply phosphorylated pHis peptides. Triple score versus ptmRS score was therefore considered for singly, doubly and multiply (≥3) phosphorylated peptides separately (Figure 5.8 b, c and d). For multiply phosphorylated peptides a triplet score of 0 is most likely; 70% of multiply phosphorylated peptides have triplet score of 0 across all ptmRS score categories. The trends for singly and doubly phosphorylated peptides are broadly similar, with higher ptmRS score not necessarily resulting in observation of the triplet pattern. For additional investigation of the triplet neutral loss pattern only singly phosphorylated peptides were considered.

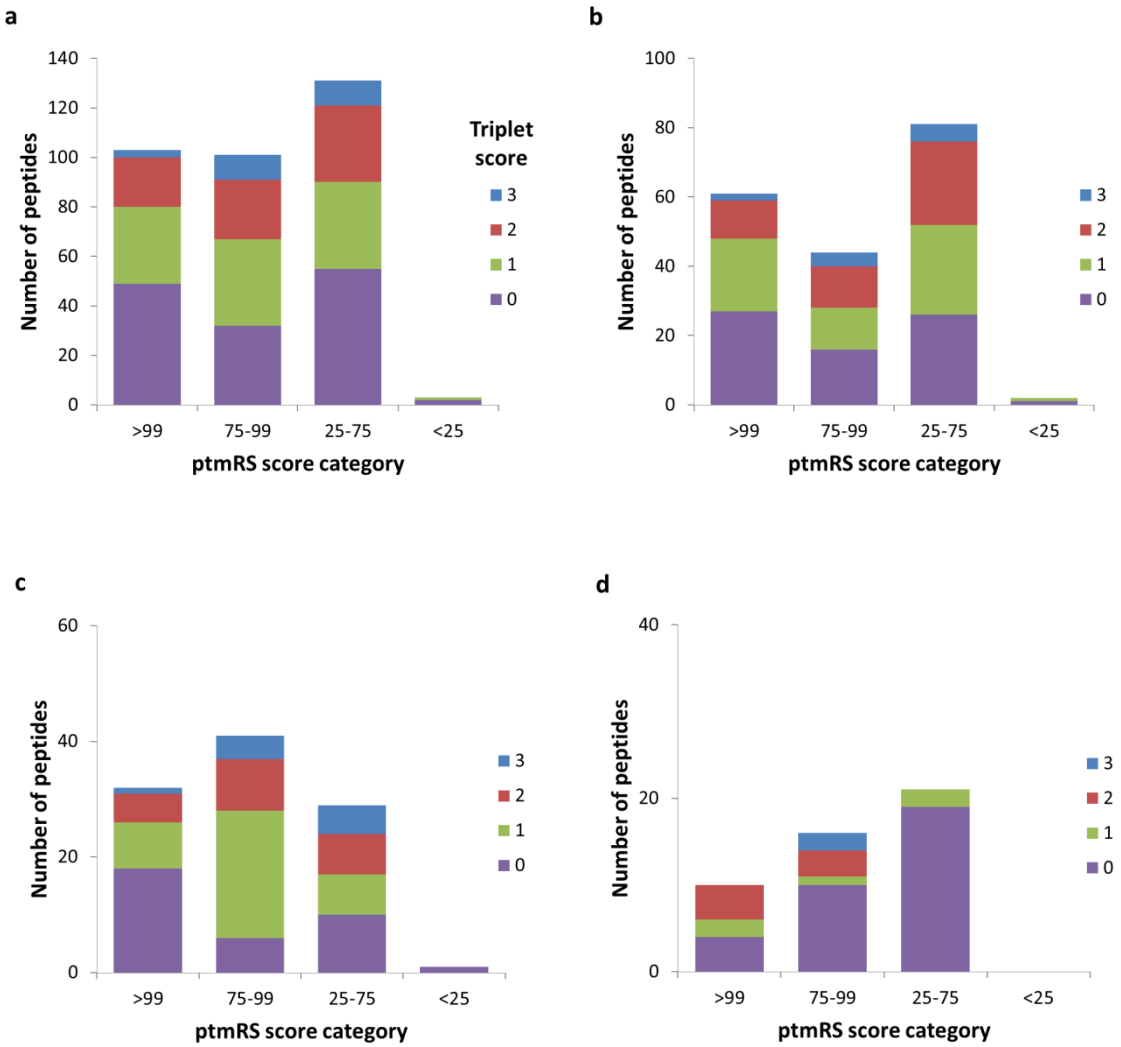


Figure 5.8 Distribution of triplet scores for pHis peptides as a function of site localisation confidence. Data shown is for unique phosphopeptides identified across 16 SAX fractions, with site confidence determined by ptmRS score and neutral loss ions considered present if ≥2% compared to the base peak ion. Triplet score of 3 indicates observation of all three neutral loss ions, 2 indicates any two neutral loss species and 1 indicates any one of the three possible neutral loss ions. **a)** all phosphopeptides containing a pHis site, **b)** singly phosphorylated pHis containing peptides, **c)** doubly phosphorylated pHis containing peptides **c)** multiply phosphorylated (≥3 phosphosites) pHis containing peptides

Is triplet score related to MASCOT score or precursor intensity?

One possible reason for the failure of these potential pHis peptides to adhere to the triplet neutral loss pattern could be a result of poorer quality spectra. As a measure of this the MASCOT score was used, whereby higher MASCOT scores may generally indicate better quality spectra, and thus these higher quality spectra may be more likely to contain the indicative triplet neutral loss pattern for pHis (Figure 5.9 a). Additionally, it was proposed that spectra arising from peptides with higher signal intensity would be more likely to display the triplet neutral loss pattern; the neutral loss ions arising from low intensity peptides would be more likely to fall below the limits of detection. The triplet scores were considered for peptides in the top 10% and top 25% for precursor signal intensity, with peptides in the bottom 25% when ranked by signal intensity also shown for comparison (Figure 5.9 b).

The highest number of pHis peptides with a triplet score of 3 are observed in the >40 MASCOT score category; 13% of pHis peptides in the >40 category have a triplet score of 3, compared to 10%, 3% and 2% in the 30-40, 20-30 and <20 categories respectively. However, this category still contains a majority of peptides with a triplet score less than 3, and 20% of these peptides display no neutral loss ions. The trend is similar for the 30-40 MASCOT score category. For peptides with a MASCOT score below 30 there is an even greater proportion of peptides (39%) with a triplet score of 0, whilst over half (57%) of the peptides with a MASCOT score below 20 have no neutral loss ions. However, MASCOT score still does not fully account for the observed triplet scores for these pHis peptides.

When considering the relationship between signal intensity and triplet score, the top 10% most intense ions still only contain a small proportion of peptides (11%) with a triplet score of 3, with almost half (42%) of these peptides again not displaying any neutral loss. The trend is similar for peptides in the top 25%, whilst the least intense peptides seem to exhibit the highest proportion of one and two neutral losses, with 64% of peptides having triplet score of 1 or 2. Therefore, the failure of these pHis peptides to adhere to the triplet neutral loss pattern does not appear to be related to signal intensity.

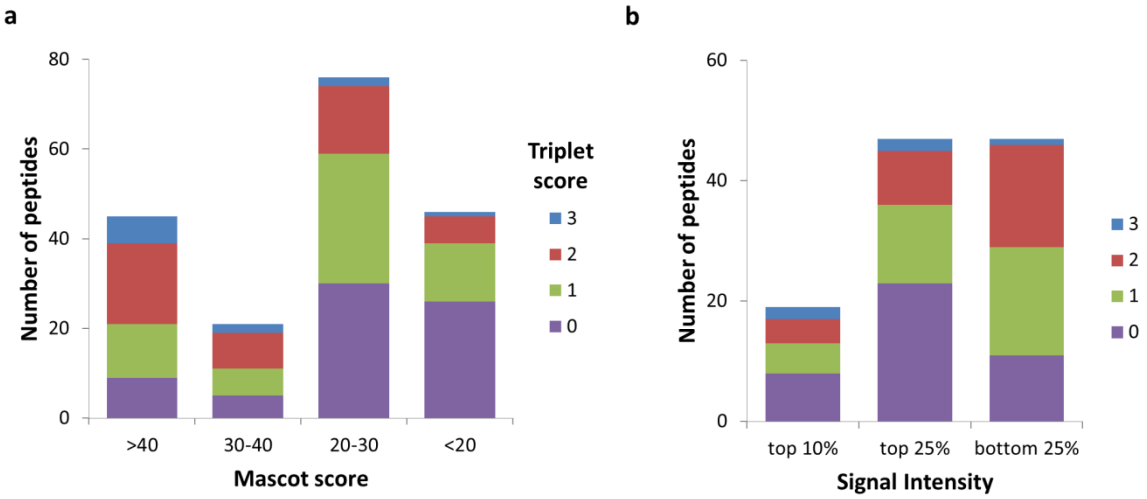


Figure 5.9 Distribution of triplet scores for pHis peptides as a function of a) MASCOT score and b) signal intensity. Data shown is for unique singly phosphorylated peptides identified across 16 SAX fractions, with neutral loss ions considered present if ≥2% compared to the base peak ion.

Do triplet scores and observed neutral loss ions follow expected trends for other sites of phosphorylation?

To establish whether the failure of pHis peptides to follow the expected neutral loss pattern was specific to these peptides, or a result of all phosphopeptides failing to exhibit the expected neutral losses, triplet scores were considered for all phosphopeptides across the 16 fractions. Based on expected neutral losses from phosphopeptides pSer/pThr should have triplet scores of 1 or 2 (relating to Δ98 Da and/or Δ116 Da) whilst pTyr peptides should exhibit only 1 neutral loss (Δ80 Da). None of these phosphopeptides should have the triplet neutral loss pattern. The triplet scores for each site of phosphorylation are shown in relation to the ptmRS score categories, with the expectation that more confidently localised sites may be more likely to exhibit the expected neutral losses. Only singly phosphorylated peptides were considered, and the number of peptides with each triplet score has been converted to a percentage to allow for easier comparison when pSer/pThr peptides far outnumber pHis and pTyr (Figure 5.10).

The expected pattern of triplet score is not clearly observed for any phosphopeptides. For all sites of phosphorylation peptides with a triplet score of 3 are observed in almost all of the score categories, with the greatest percentage of peptides displaying the triplet pattern actually observed for pSer/pThr peptides. In the >99 score category pSer/pThr peptides do predominantly have scores of 1 or 2 as would be expected, whilst close to 40% of pTyr peptides in this category have a score of 1, again in line with the expected neutral loss. These trends are lightly less clear as the ptmRS score decreases, and many peptides regardless of site of phosphorylation or ptmRS score have no neutral loss ions (triplet score of 0).

The data was further interrogated to determine the actual neutral loss ions (Δ80, Δ98 and/or Δ116 Da) rather than an overall triplet score, with the aim of revealing more clearly whether the expected mechanisms of neutral loss are occurring for each site of phosphorylation. For this analysis only phosphopeptides with ptmRS score ≥ 75 were considered, as there is a greater confidence in the assignment of these phosphosites so the neutral loss ions should match the proposed site (Figure 5.11).

For pSer/pThr peptides with a triplet score of 1, the most common neutral loss is Δ98 Da, accounting for 54% of the peptides in this category. For a triplet score of 2 the predominant

combination of neutral losses, corresponding to 74% of peptides in this category, is Δ98 Da and Δ116 Da, suggesting these peptides are following the expected neutral loss behaviour to some degree. For pTyr 41% of the peptides have a triplet score of 1, as would be expected, but the type of neutral loss observed is fairly evenly split across Δ80, Δ98 and Δ116 Da. This even spilt of neutral loss ions is also observed for pHis peptides.

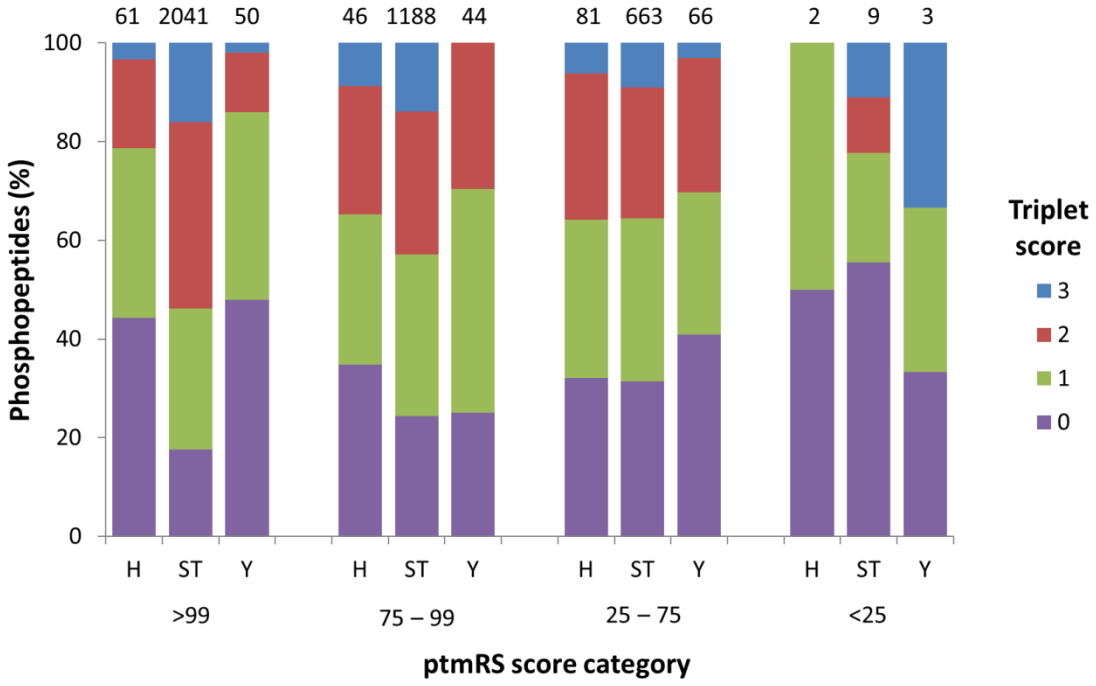


Figure 5.10 Distribution of triplet scores for singly phosphorylated peptides (pHis, pSer/pThr and pTyr) as a function of site localisation confidence. Data shown is for unique phosphopeptides identified across 16 SAX fractions, with site confidence determined by ptmRS score and neutral loss ions considered present if ≥2% compared to the base peak ion.

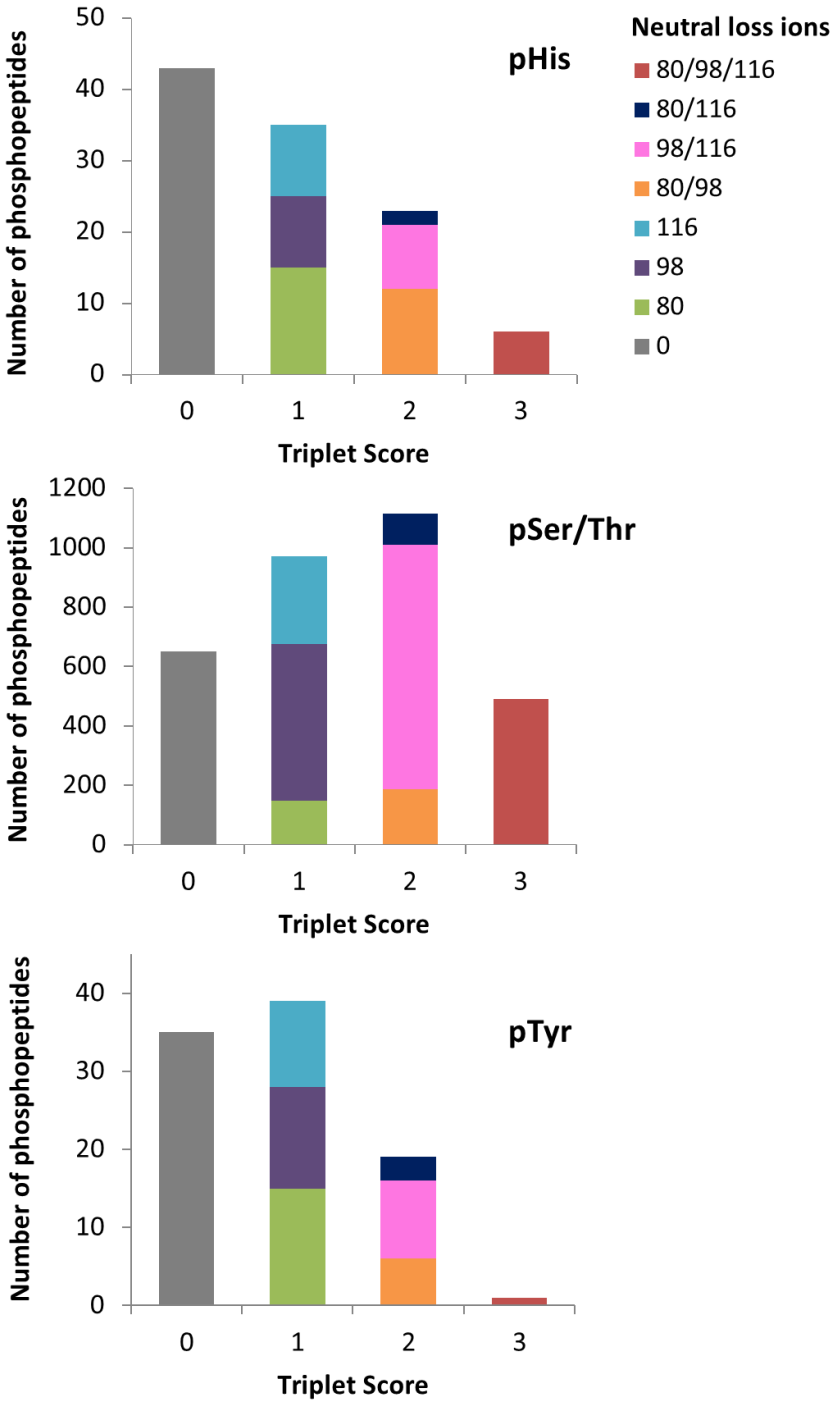


Figure 5.11 Number of phosphopeptides exhibiting different combinations of neutral loss ions (Δ80, Δ98 and/or Δ116 Da). Neutral loss ions identified at ≥2% compared to base peak signal intensity for **a)** pHis, **b)** pSer/pThr, **c)** pTyr. Data is shown for unique singly phosphorylated peptides where the site localisation confident as determined by ptmRS is ≥75.

Does relative intensity of neutral loss ions affect triplet scoring?

Up to this point a neutral loss ion has been considered to be present if its intensity is ≥2% compared to the base peak ion in the spectrum. This threshold is fairly low, meaning that peaks which are just noise may be assigned as neutral loss ions, potentially explaining why pSer, pThr and pTyr peptides are observed to have more neutral loss ions than expected. Therefore, the threshold for determining a neutral loss ion was adjusted to 5% and 10% compared to the base peak intensity, then the number and nature of the neutral loss ions for each peptide was again determined. For the data with a 5% relative signal intensity cut-off, the triplet score was also determined for non-phosphorylated peptides. As these peptides cannot undergo neutral loss (due to the absence of a phosphate group) any ions corresponding the expected *m/z* of a neutral loss are just background noise in the spectrum, thus providing an idea of the degree of false positives when considering neutral losses in this way (Figure 5.12).

For pSer/pThr phosphopeptides increasing the intensity cut-off makes the expected trend of predominantly Δ98 Da and/or Δ116 Da neutral losses much more apparent. At a 10% cut off only 2% of pSer/pThr peptides exhibit the triplet neutral loss, and only a further 4% have a Δ80 Da neutral loss either alone or with another neutral loss ion. For pTyr peptides the number of Δ116 Da neutral loss ions reduces as the cut off is increased, meaning that Δ80 Da and Δ98 Da are the most common neutral loss ions observed; 23% of pTyr peptides exhibit one or both of these neutral losses when the 5% threshold is applied. Although the phosphorylated tyrosine itself should only give rise to a neutral loss of 80 Da, it appears that loss of water from elsewhere in the peptide results in Δ98 Da also being observed, with the number of peptides displaying one of these two neutral losses approximately equal when a 10% signal intensity cut-off is used. The observed neutral loss trends for pTyr are broadly similar to pHis peptides, meaning the mechanisms by which neutral loss occurs from these peptides is probably also similar. However, caution must be taken in drawing conclusions from these neutral loss patterns, as ions corresponding to neutral loss masses are also observed in the spectra of non-phosphorylated peptides; 26% of all non-phosphorylated peptides display at least one 'neutral loss' ion when the 5% intensity threshold is applied. The proportion of 'neutral loss' containing spectra for non-phosphorylated peptides compared to pHis and pTyr phosphopeptides is fairly similar (38% of pHis and 35% of pTyr peptide spectra contain at least one neutral loss peak), meaning

there could be a considerable number of false positives, masking the true trends in this data.

One very clear observation in this data is the number of singly phosphorylated peptides (at 5% FDR with ptmRS score ≥75) that do not exhibit any neutral loss ions (triplet score of 0). Although neutral loss from phosphopeptides is widely reported to occur following CID fragmentation, these neutral loss ions do not appear to be a dominant feature in HCD spectra. Whilst this may mean that the HCD spectra being generated are more useful for phosphopeptide identification and site localisation, by virtue of more complete backbone fragmentation which can otherwise be hindered by extensive neutral loss, it also means that using neutral loss ions as an additional tool to distinguish sites of His phosphorylation may be an unsuccessful strategy.

Overall, it appears that the triplet neutral loss pattern cannot be used as a characteristic marker of pHis peptides. A number of high scoring and confidently localised pHis peptides (according to ptmRS score) did not display the expected triplet of neutral loss ions. Manual inspection of a few example spectra confirms that the absence of the triplet neutral loss is not detrimental to the identification of pHis peptides; confident site localisation can be achieved despite a triplet score of zero or one (Figure 5.13). For future phosphopeptide analysis the triplet pattern was not considered, instead ptmRS scores were used as a measure of phosphosite confidence, with a 1% false localisation rate (FLR) applied. Based on work by Ferries *et al*.[200] appropriate score cut-off values were chosen: 99.2 for HCD data and 90.2 for EThcD, reflecting the greater confidence in site localisation afforded by EThcD fragmentation.

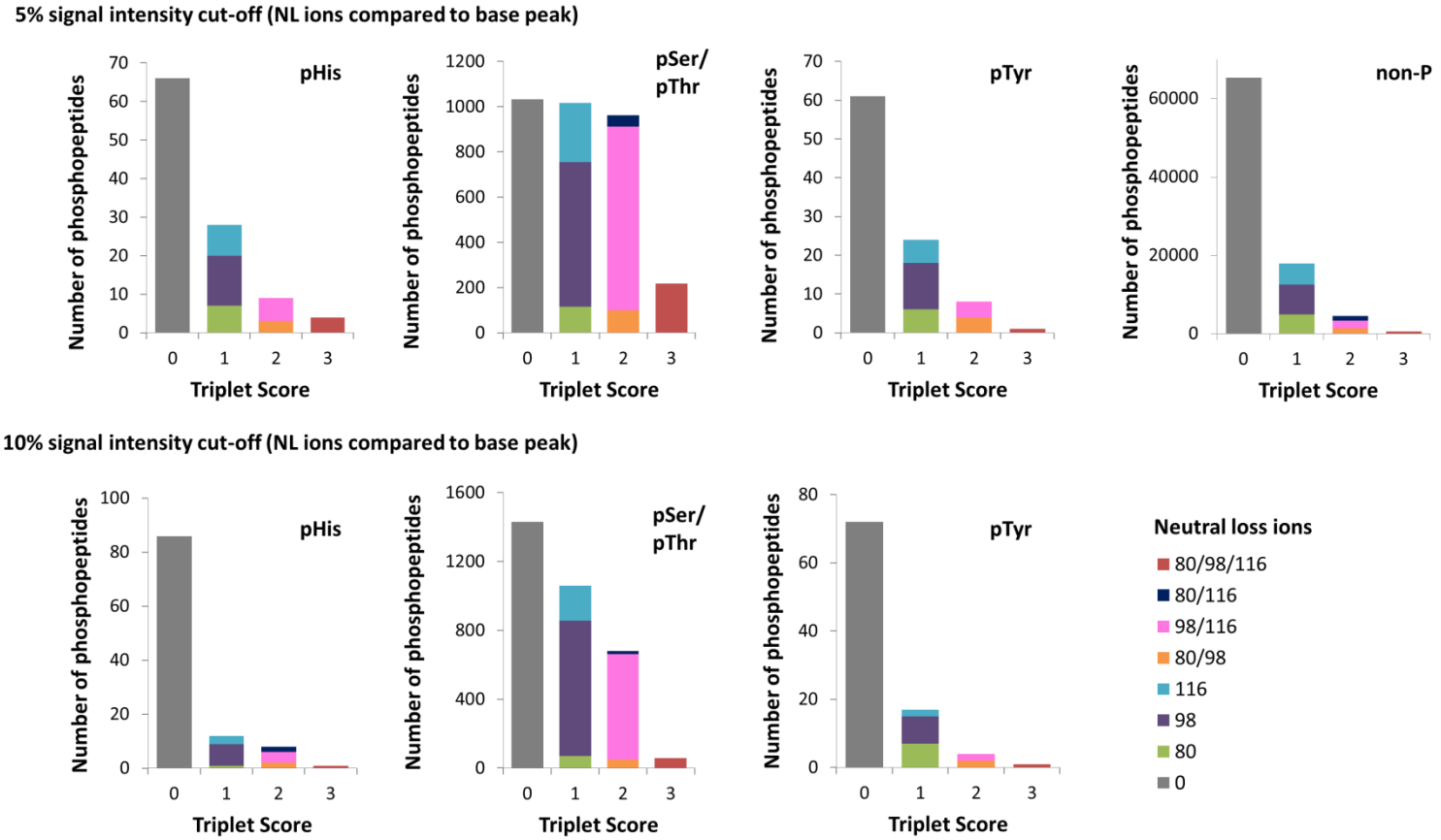


Figure 5.12 Number of phosphopeptides exhibiting different combinations of neutral loss ions (Δ80, Δ98 and/or Δ116 Da) at 5% or 10% signal intensity cut-off a-d) Neutral loss ions identified at ≥5% compared to base peak signal intensity for **a)** pHis, **b)** pSer/pThr, **c)** pTyr and **d)** non-phosphorylated peptides, **e-g)** Neutral loss ions identified at ≥10% compared to base peak signal intensity for **e)** pHis, **f)** pSer/pTh and **g)** pTyr**.** Data is shown for unique singly phosphorylated peptides where the site localisation confidence as determined by ptmRS is ≥75.

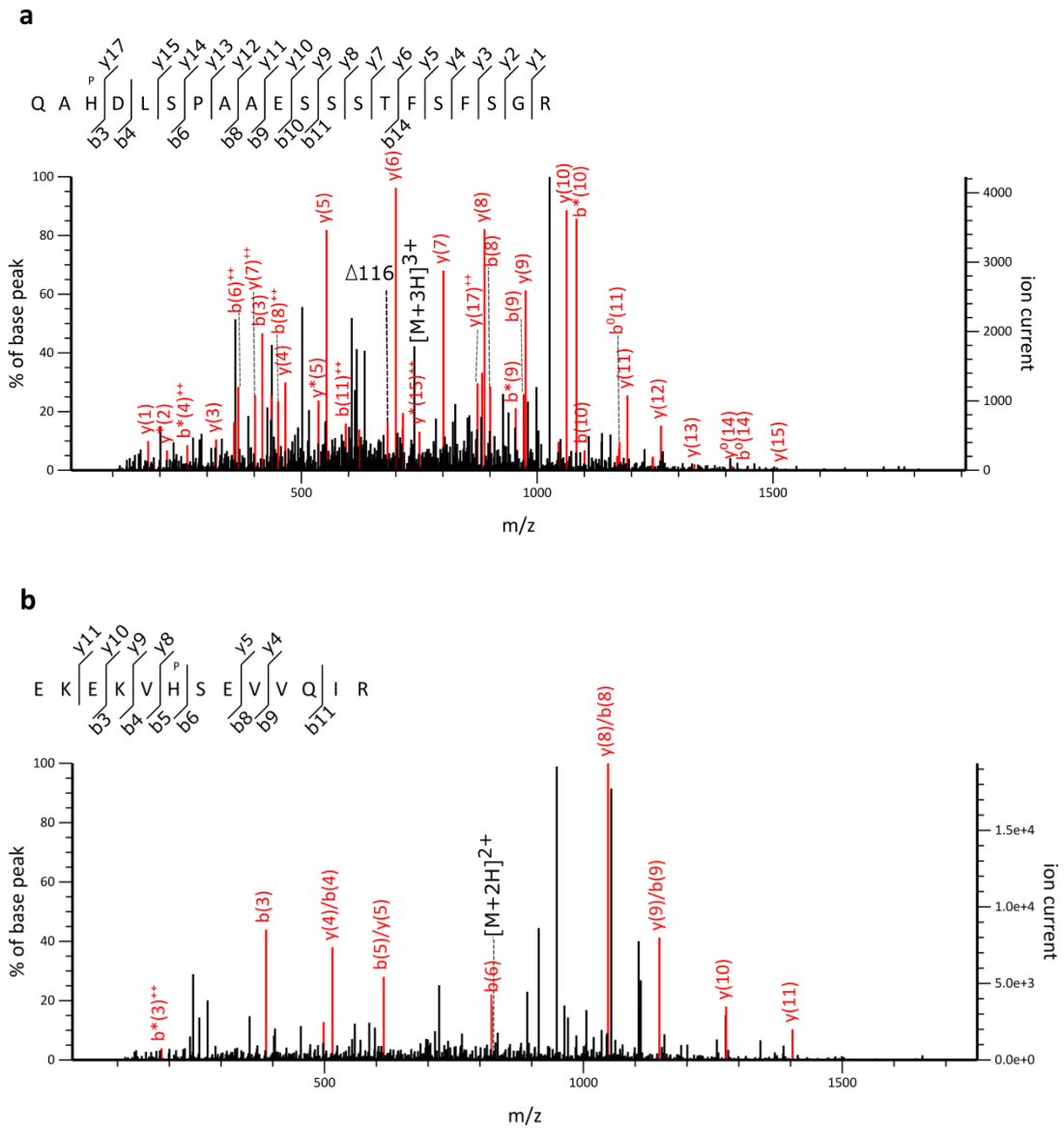


Figure 5.13 Example pHis spectra demonstrate confident pHis localisation without triplet neutral loss a) HCD fragmentation of triply charged peptide ion QApHDLSPAAESSSTFSFSGR (*m/z* 721.31), exhibiting only one neutral loss ion (Δ116 Da), with ptmRS score H3:99.79 **b)** HCD fragmentation of doubly charged peptide ion EKEKVpHSEVVQIR (*m/z* 830.93), exhibiting no neutral loss ions, with ptmRS score H6:99.92.

5.2.5 Phosphoserine, threonine and tyrosine proteins

In order to assess the unbiased nature of the SAX method presented in this work, the identified S/T/Y-phosphoproteins were compared to the phosphoproteins identified by an alternative enrichment strategy reported in the literature (Figure 5.14). The work by Sharma *et al.*[251] describes the phosphoproteins identified from HeLa lysates that were first fractionated by SCX then enriched for phosphopeptides using TiO_2. This two-stage fractionation and enrichment approach employed for deep phosphoproteome coverage results in a greater number of phosphoprotein IDs in the Sharma *et al.* dataset; 5545 proteins compared to 2080.

The proteins chosen for comparison were those identified in the control condition containing at least one 'Class I' S/T/Y-phosphosite, that is with a PTM Score >75 according to the MaxQuant phosphosite localisation algorithm. It is worth noting that the MaxQaunt software utilises the Andromeda search engine, whereas this work used MASCOT with ptmRS for phosphosite localisation, which may result in differences between identified phopshopeptides[200]. Nevertheless, the majority (77%) of S/T/Y phosphoproteins identified by this SAX method (and the associated data analysis pipeline) are also present in the Sharma *et al.* dataset. This high degree of overlap indicates that the SAX method recovers an unbiased sub section of the phosphoproteome.

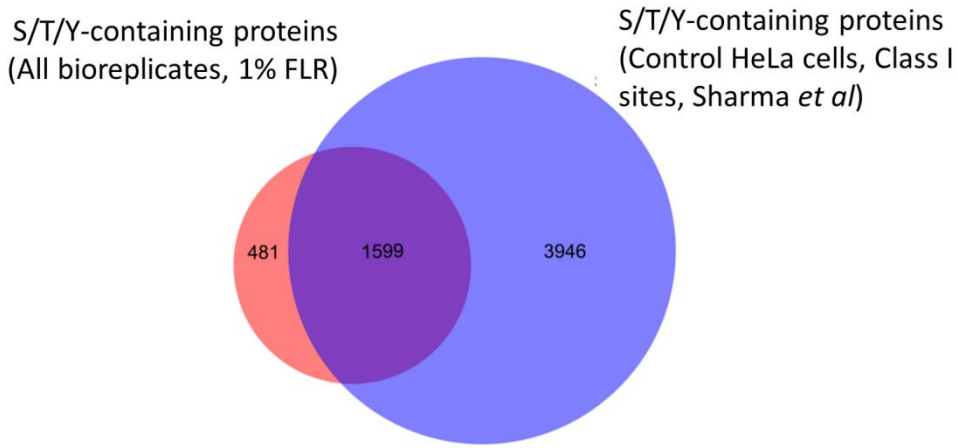


Figure 5.14 Comparison of S/T/Y-phosphoprotein IDs in HeLa cells. 77% of the S/T/Y phosphoproteins identified by the method presented in this work (with at least one S/T/Y-phosphosite above 1% FLR) are also present in the Sharma *et al.* dataset (phosphoproteins with Class I phosphosites (PTM score > 75) identified following SCX fractionation and TiO_2 enrichment[251]).

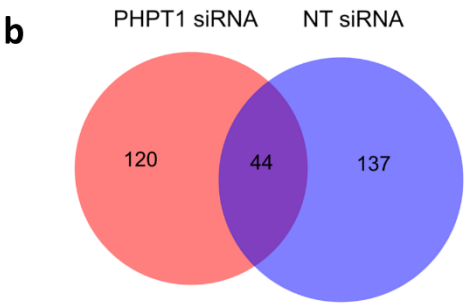
5.2.6 Phosphohistidine-containing proteins

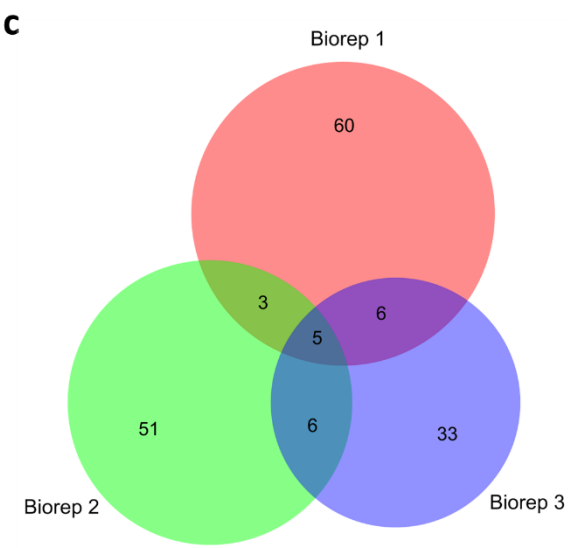
Application of the previously described data analysis pipeline results in a total of 301 pHis peptides identified above the 1% FLR cut-off. 164 unique pHis peptides (corresponding to 172 pHis sites) were identified in the PHPT1 siRNA samples, compared to 181 pHis peptides (186 pHis sites) in the NT siRNA sample (Figure 5.15). In both cases overlap between bioreplicates was fairly low with 88% and 84% of peptides identified in only one bioreplicate for PHPT1 siRNA and NT siRNA conditions respectively. Of the 164 peptides in the PHPT1 siRNA condition 27% (44 peptides) are also seen in the NT siRNA condition. Contrary to what was expected the number of pHis peptides is higher in the NT siRNA condition compared to PHPT1 siRNA treated samples. It was hypothesised that knockdown of a His phosphatase would increase levels of pHis. However, it is possible that some degree of negative feedback is occurring resulting in decreased pHis levels in the absence of PHPT1. Additionally, PHPT1 is not the only known His phosphatase[35] and it is possible that here is redundancy in terms of their substrates. Indeed, it is also possible that other as yet unknown phosphatases could be acting to reduce overall pHis levels compared to the control sample.

Analysis of PHPT1 siRNA and NT siRNA treated HeLa cells using the UPAX method has resulted in the identification of 296 pHis-containing proteins, which is largest number of pHis proteins identified in a single study where the site of phosphorylation is known. Only a small number of proteins identified in this work have previously been identified as phosphorylated on His. Interestingly, pHis peptides from well characterised pHis proteins such as histones and the NME1/2 kinases have not been identified, suggesting this is a far from complete analysis of the pHis proteome. A selection of pHis proteins/peptides are shown in Table 5.1 (the full list can be found in the Appendix).

a

	PHPT1 siRNA	NT siRNA
Unique pHis peptides	164	181
Unique pHis sites	172	186
Unique pHis proteins	164	176





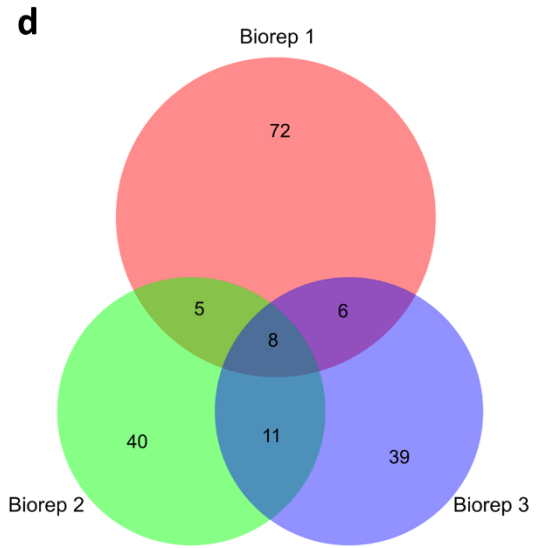


Figure 5.15 Number of pHis peptides identified in PHPT1 siRNA and NT siRNA conditions with 1% FLR based on ptmRS localisation score. a) Number of unique peptides, sites and proteins identified in each condition. **b)** overlap in pHis peptide IDs between PHPT1 siRNA and NT siRNA conditions **c)** overlap in pHis peptides IDs between bioreplicates for PHPT1 siRNA **d)** overlap in pHis peptide IDs between bioreplicates for NT siRNA

Table 5.1 A selection of pHis peptide/protein IDs that were observed at least twice across 3 bioreplicates; the highest scoring ID is shown. Localised pHis residue is indicated as 'pH', other phosphorylation sites in the peptide not indicated (further details in Appendix).

Protein Description	Protein Accession	Peptide sequence	Site (protein)	ptmRS score (pHis site)	Ion score	Condition
26S proteasome non-ATPase regulatory subunit 14[a]	O00487	MLLNLpHK	221	100	25	Both
ARF GTPase-activating protein GIT1	Q9Y2X7	pHGSGADSDYENTQSGDPLLGLEGK	590	99.89	53	Control
CAD protein[a]	P27708	MVVMpHPMPR	2183	100	24	PHPT1
Calumenin	O43852	VpHNDAQSFDYDHDAFLGAEEAK	39	99.72	76	Both
Cyclin-dependent kinase 19	Q9BWU1	ELKpHPNVIALQK	75	100	37	Both
Cytoplasmic dynein 1 heavy chain 1[a]	Q14204	KLMpHVAYEEFEK	403	100	20	Control
E3 ubiquitin-protein ligase CHIP	Q9UNE7	KDIEEpHLQR	260	100	22	PHPT1
Elongator complex protein 1	O95163	TPpHLEKRYK	481	99.72	21	Control
FACT complex subunit SSRP1[a]	Q08945	EGMNPSYDEYADSDEDQpHDAYLER	449	99.44	45	PHPT1
Golgin subfamily A member 4[a]	Q13439	LQKLpHEKELAR	485	100	19	Both
PHD finger protein 6[a]	Q8IWS0	TAHNSEADLEESFNEpHELEPSSPK	149	99.32	29	Control
Pre-mRNA 3'-end-processing factor FIP1	Q6UN15	ERDpHSPTPSVFNSDEER	491	99.23	78	Both
Protein RRP5 homolog[a]	Q14690	ALECLPSKEpHVDVIAK	1760	99.89	30	PHPT1
Ribosomal L1 domain-containing protein 1[a]	O76021	EQTPEpHGKK	343	100	28	Both
RNA polymerase II elongation factor ELL2	O00472	VIpHLLALK	212	100	30	Both
Serine/threonine-protein kinase PLK2	Q9NYY3	EIELpHRILHHK	134	100	26	Both
T-box transcription factor TBX4	P57082	KFpHEAGTEMIITK	81	100	33	PHPT1
T-complex protein 1 subunit alpha[a]	P17987	DDKpHGSYEDAVHSGALND	542	99.94	46	Both
Ubiquitin carboxyl-terminal hydrolase 43	Q70EL4	LSTLVKFPLSGLNMAPpHVAQR	618	100	22	Control

a. identified as pHis proteins by Fuhs *et al.*[227]

5.2.7 Enrichment analysis and phosphohistidine motifs

In order to reveal possible biological processes related to these pHis proteins, functional clustering and enrichment analysis was performed using DAVID[259, 260] (version 6.8). The pHis proteins identified in each condition were compared to a background of all proteins identified in the samples to determine features that are enriched (Figure 5.16; full output in Appendix). For each cluster, the fold enrichment and associated P-value is shown. For this analysis only clusters containing >3 proteins were considered, with low protein counts tending to suggest weaker evidence for enrichment. DAVID calculates P-values using a modified Fisher's exact test, also called the EASE score, which is a more conservative method than a standard Fisher's exact test. A Benjamini P-value is also calculated, which is a multiple testing correction technique used to globally correct P-values.

Proteins in the NT siRNA condition are significantly enriched for motor protein (8 proteins, 5.7 fold, P=4.58E-4) with cilium (7 proteins, 4.3 fold, P=5.43E-3), zinc finger, C2H2 (5 proteins, 5 fold, P=1.75E-2) and heme binding (4 proteins, 6.9 fold, P=1.94E-3) also enriched. Upon treatment with PHPT1 siRNA only the zinc finger, C2H2 cluster is still found to be enriched amongst the identified proteins (6 proteins, 4.7 fold, P=8.57E-3). Enrichment is also reported for transducer (7 proteins, 4.6 fold, P=3.81E-3) and kinase activity (7 proteins, 3.5 fold P=1.44E-2). There were no enriched KEGG pathways identified for either dataset.

Analysis with DAVID incorporates terms from a variety of sources, utilising over 40 annotation categories, including GO terms, protein functional domains and sequence features. This extended annotation coverage provides a comprehensive view of the data. Potential limitations of functional annotation and enrichment tools, such as DAVID, include incomplete annotation databases and the fact that only existing annotations are included, i.e. such informatics methods cannot discover previously unknown functions of any given protein cohort[261]. It is thus theoretically possible that the pHis-mediated regulation of any given protein identified in this study could be a moonlighting function in addition to the function already annotated and described in the various databases.

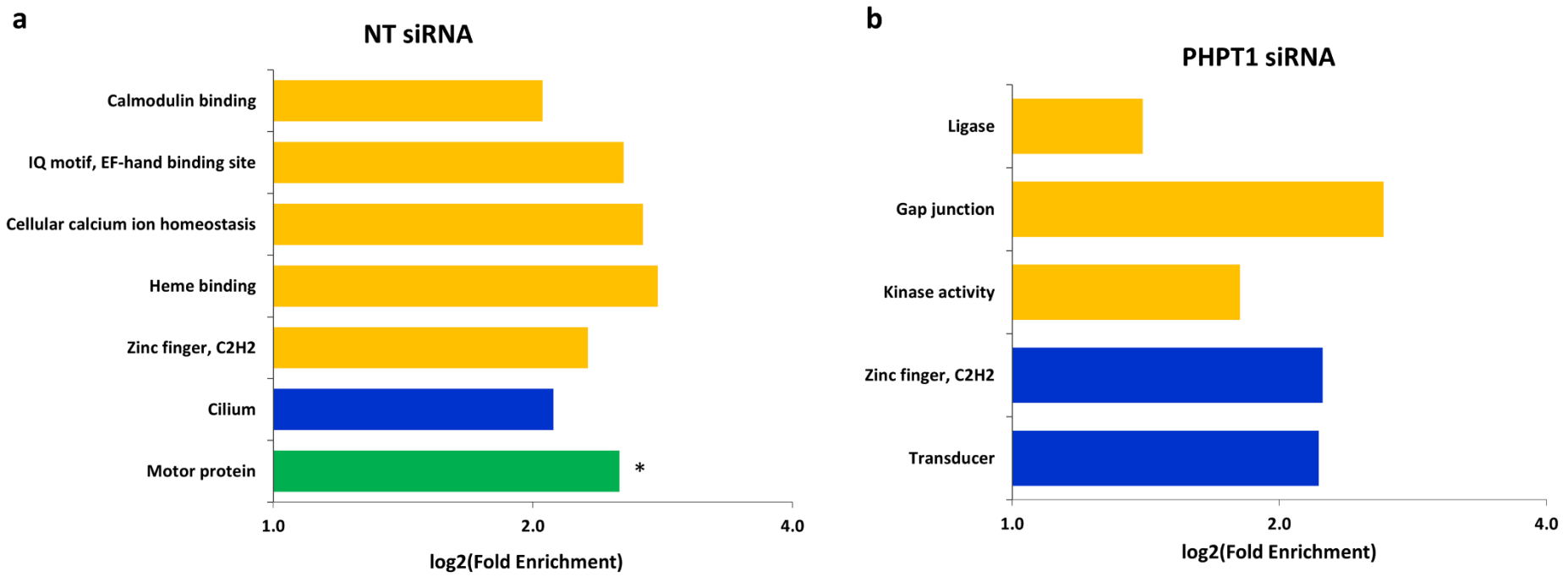


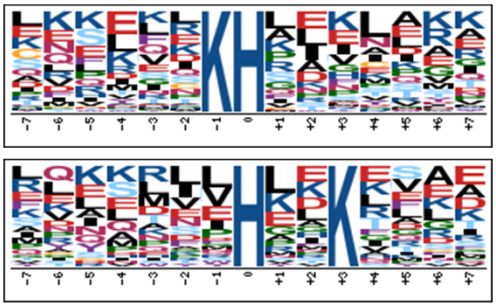
Figure 5.16 Enrichment analysis of pHis proteins performed using DAVID. **a)** proteins identified in NT siRNA condition **(b)** proteins identified in PHPT1 siRNA condition. P-value indicated by colour: yellow P<0.05, blue P<0.01, green P<0.001 with asterisks defining clusters where the Benjamini adjusted P-value <0.1.

Given that this dataset reports the largest number of identified pHis peptides/sites to date, this presents a valuable opportunity to examine motifs surrounding these sites of His phosphorylation. The seven residues either side of a confidently localised pHis site (1% FLR) were used to search for pHis motifs using the Motif-X tool[262, 263] (Figure 5.17). A search with all pHis sites revealed two distinct motifs, both of which were Lys directed. Lys is enriched at either the -1 (2.84 fold) or +3 position (2.63 fold) in these motifs. Additionally, there appears to be a preference for Lys, Glu and Leu in the nearby residues either side of a phosphorylated His. Given the charge directed nature of the SAX separation it was suspected that there could be a bias in the motifs observed in the early compared to the later fractions. Motifs for fractions 1-8 are again all Lys directed: -1 position (3.9 fold), +3 position (3.62 fold) and -3 position (3.90 fold), whilst the later eluting fractions (9-16) exhibit a preference for Glu at +5 (3.17 fold).

a

All fractions

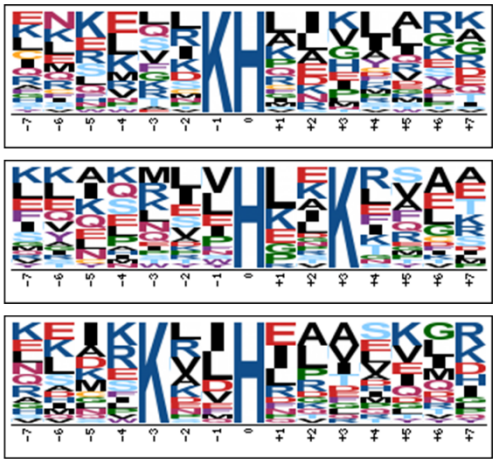
Motif	Score	Matches in dataset	Fold increase
……KH…….	9.21	44	2.84
…….H..K….	6.47	34	2.63



b

Fractions 1-8

Motif	Score	Matches in dataset	Fold increase
……KH…….	7.04	24	3.49
…….H..K….	6.24	20	3.62
….K..H…….	6.17	18	3.90



c

Fractions 9-16

Motif	Score	Matches in dataset	Fold increase
…….H….E..	7.71	30	3.17

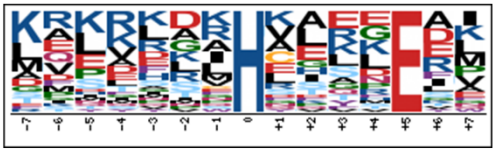


Figure 5.17 Extracted motifs from pHis peptide data. a) Motifs from pHis peptides in all 16 SAX fractions **b)** Motifs from pHis peptides in SAX fractions 1-8 **c)** Motif from pHis peptides in SAX fractions 9-16. A higher score indicates a more significant/more specific motif. Fold increase compared to background of all sequences in the human proteome. Significance set at 1E-5 (not corrected for multiple hypotheses, corresponds to P value of 3E-4 with Bonferroni correction).

5.2.8 A potential role for phosphohistidine in kinases

Reversible phosphorylation of proteins by protein kinases is a vital mechanism for eukaryotic signalling. Of the ~300 pHis peptides identified in this analysis, 12 were from protein kinases, which represents ~4% of all pHis peptide identifications. A sequence alignment (performed by P. Eyers, University of Liverpool) revealed that three of these pHis sites map to the identical region of the protein kinase catalytic domain, specifically the HxN-His position of the αC-β4 loop. Furthermore, the pHis site is conserved in >65% of human protein kinases. Annotated spectra for the peptides derived from Ser/Thr kinases CDK19 and PLK2 and the tyrosine kinase Ephrin A1 are shown in Figure 5.18.

For both CDK19 and PLK2 multiple instances of the same pHis peptide were identified across bioreplicates and across the two conditions. The highest scoring identification for each of these peptides is displayed in Table 5.1, with further details in the Appendix. The PLK2 peptide EIELHRILHHK is doubly phosphorylated with one phosphorylation confidently localised to H5 (above the 1% FLR threshold) whilst the other phosphorylation could be on one of a further two possible His residues in the peptide (H9, H10). The conserved residue maps to H10 in the peptide, although unfortunately there is no definitive evidence to localise the phosphorylation to this residue over the adjacent H9. The CDK19 peptide ELKHPNVIALQK is singly phosphorylated; the site of phosphorylation is localised to the H4 residue with a ptmRS score of 100. The Ephrin A1 peptide EATIMGQFSHPHILHLEGVVTKR contains four sites of phosphorylation. The pHis site which maps to the conserved residue (H10) has a ptmRS score of 96.87, which is below the 1% FLR threshold but still above the cut-off of 75 many phosphoproteomics studies use to confidently assign phosphosites.

The αC-β4 loop, to which these pHis sites map, is implicated in positioning of the αC-helix and Hsp90 mediated kinase maturation[264-266]. Additionally, PTMs of the αC-β4 loop have been observed in multiple kinases and are proposed to regulate the position of the preceding αC-helix[267, 268]. Therefore, phosphorylation of the HxN-His residue might act as a general mechanism for regulating kinase activity by altering kinase dynamics. To further investigate this hypothesis, molecular dynamics (MD) simulations have been performed (by N. Kannan & Z. Ruan, University of Georgia) (Figure 5.19). They used PLK2 as the model system; His139 maps to the conserved HxN-His position and corresponds to H10 in the observed peptide sequence (EIELHRILHHK). MD simulations were carried out on four possible pHis139 states, based on the protonation state and the attachment of the

phosphate group on the His residue. Specifically, pHis139-N1-H1D is the 1-pHis isomer with a singly protonated phosphate group, pHis139-N1-H2D is the 1-pHis isomer with an unprotonated phosphate group, pHis139-N3-H1E is the 3-pHis isomer with a singly protonated phosphate group, and pHis139-N3-H2E is the 3-pHis isomer with an unprotonated phosphate group. Comparison of these simulations showed increased flexibility in the phosphorylated compared to the unphosphorylated form (Figure 5.19 b). In particular the 3-pHis (N3) form displays increased flexibility in the activation segment, the regulatory αC-helix and the catalytic HRD motif (Figure 5.19 b, c), whilst the regulatory spine residue His134 also undergoes a conformational change upon phosphorylation of the His139 residue at the 3-position (Figure 5.19 d).

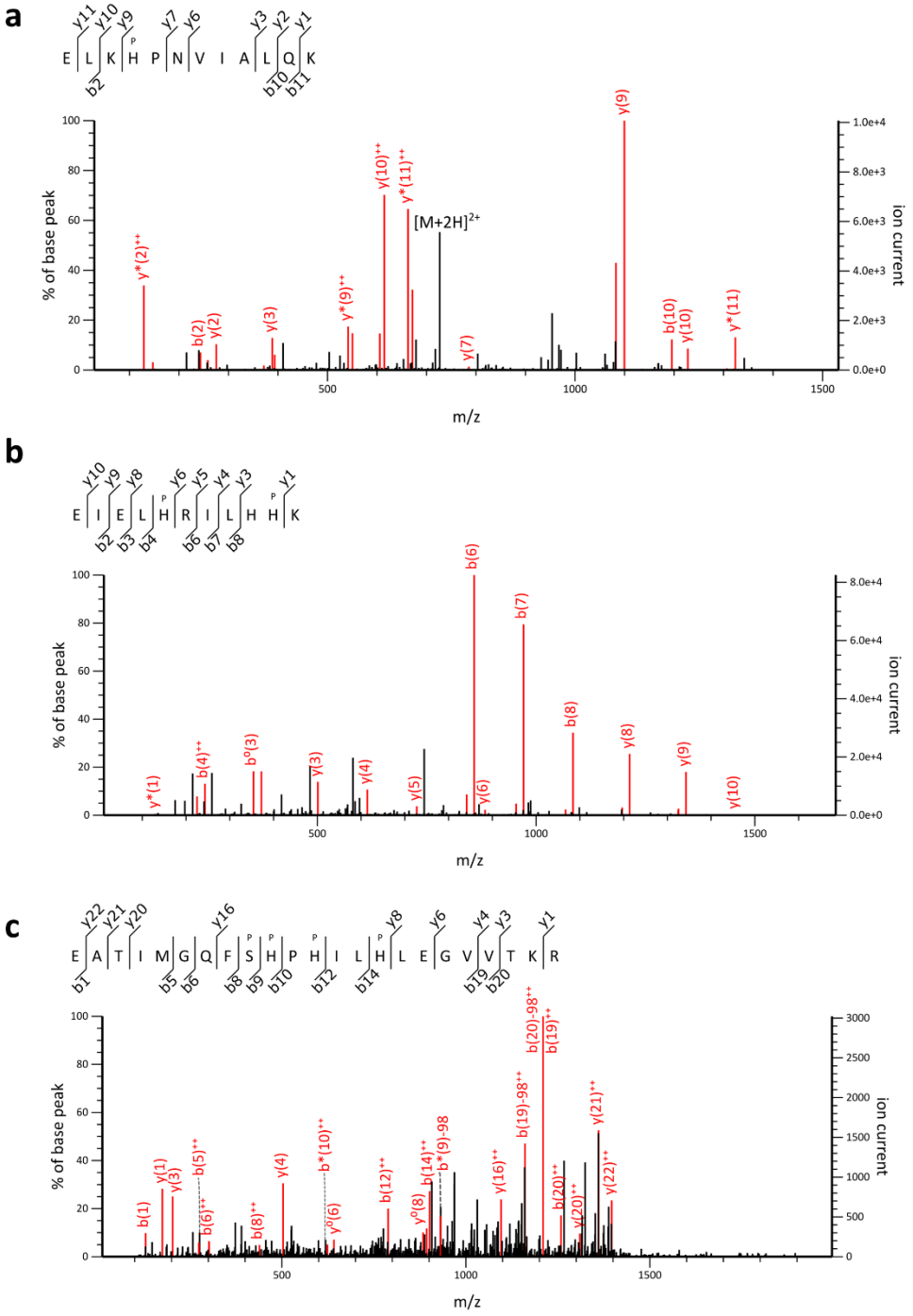


Figure 5.18 Annotated spectra for pHis peptides that map to HxN motif of protein kinases **a)** CDK19 peptide ELKpHPNVIALQK (*m/z* 735.40, ptmRS score H4:100) **b)** PLK2 peptide EIELpHRILHpHK (*m/z* 729.89, ptmRS scores H5:100, H9:50, H10:50) **c)** Ephrin A1 peptide EATIMGQFpSpHPpHILpHLEGVVTKR (*m/z* 974.09, ptmRS scores S9: 96.87, H10: 96.87, H12: 76.11, H15: 99.79)

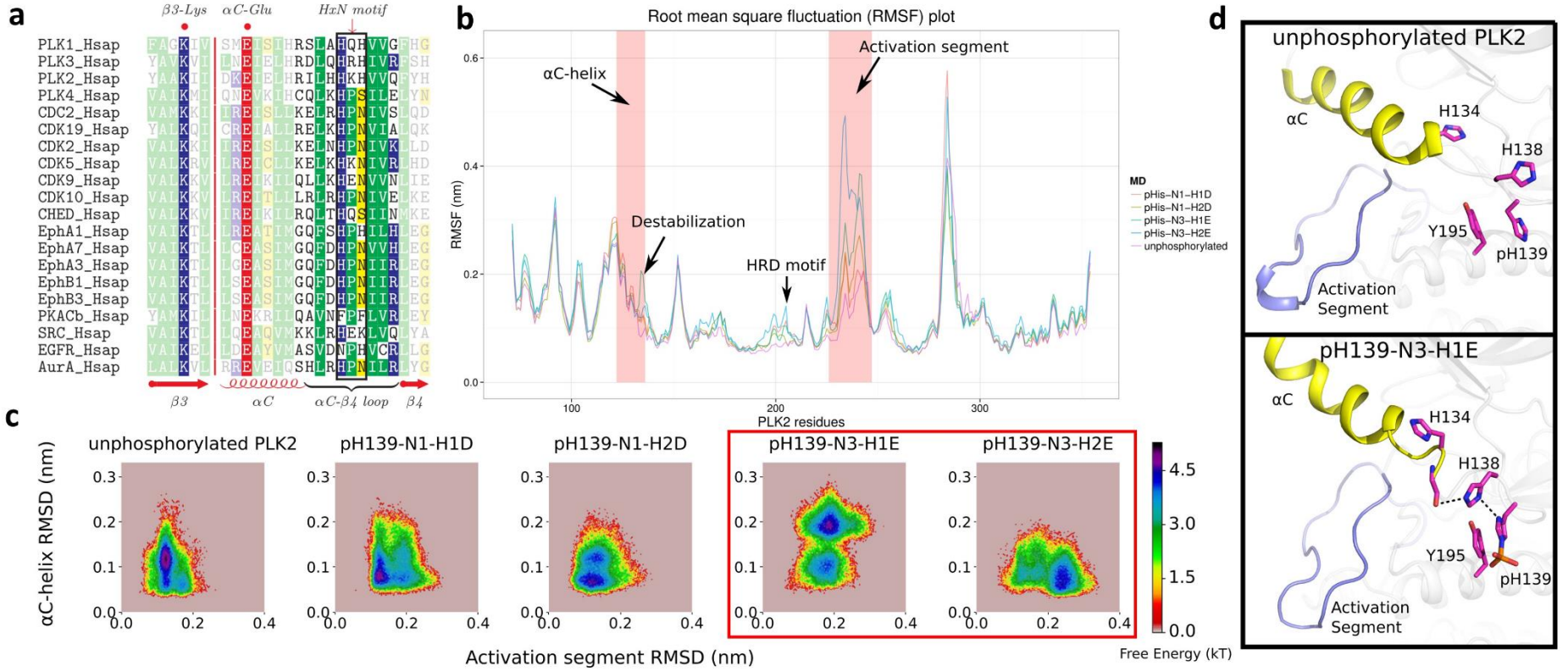


Figure 5.19 Molecular dynamics simulations of PLK2 with H139 phosphorylation a) Alignment of selected human kinases highlighting the HxN motif **b)** Root mean square fluctuation (RMSF) plot of unphosphorylated PLK2, pHis139-N1-H1D, pHis139-N1-H2D, pHis139-N3-H1E and pHis139-N3-H2E (see main text for explanation of pHis139 states). Key regions that display differential fluctuations are shown. **c)** Free energy landscape through root mean square deviation (RMSD) of αC-helix and the activation segment. **d)** Representative snapshot of unphosphorylated PLK2 (upper panel) and pHis139-N3-H1E (lower panel).

5.3 Conclusions

A novel SAX strategy for unbiased analysis of phosphopeptides has been applied to HeLa cell lysates before and after siRNA knockdown of PHPT1, a known His phosphatase. This method enables partial separation of phosphorylated from non-phosphorylated peptides, resulting in SAX fractions that are considerably enriched for phosphopeptides, an effect which was observed in all three bioreplicates across both experimental conditions.

Given the challenges associated with phosphosite localisation, an attempt was made to improve identification of pHis peptides based on a characteristic pattern of neutral loss from the precursor ion. However, the strategy did not prove diagnostic for pHis peptides. The triplet neutral loss pattern was not consistently observed for peptides where the site of phosphorylation was confidently assigned as His by the ptmRS site localisation tool and/or manual annotation of spectra. Additionally, many phosphopeptides modified on residues other than His exhibited the triplet neutral loss pattern, whilst a vast proportion of phosphorylated peptides did not show any neutral loss under the HCD fragmentation conditions employed in this experiment. The data analysis pipeline implemented for characterisation of pHis peptides therefore used ptmRS for site localisation, with a score threshold applied to achieve 1% FLR.

Characterisation of phosphopeptides in both the NT siRNA and PHPT1 siRNA conditions (across all three bioreplicates) revealed a total of 301 pHis peptides, corresponding to 314 unique pHis sites localised above the 1% FLR threshold. It was anticipated that PHPT1 siRNA treatment would increase the levels of His phosphorylation compared to the NT siRNA control. However, an increase in pHis was not observed under these conditions. Given the widespread differences between the pHis proteins identified in the two conditions it is likely that complex signalling mechanisms are involved to bring about the phosphorylation/dephosphorylation of many of these proteins, as opposed to them being direct targets for dephosphorylation by PHPT1. Hence over the time scale of the siRNA treatment more extensive effects are observed than just a rapid build-up of the pHis phosphorylation that is targeted by PHPT1 phosphatase activity.

Functional analysis of the pHis proteins identified in each of the two conditions did not identify any clear functions or localisation to a particular sub-cellular compartment, which

may suggest pHis has a more generalised role in human cells. Additionally, it should be noted that many previously identified pHis-containing proteins were not observed in this analysis, which suggests the coverage of the pHis proteome represented by this dataset is incomplete. For example, peptides from neither NME1/2 (known His kinases) or histone H4 were identified. Furthermore, there were few similarities between the pHis proteins identified in this dataset compared to the previously published pHis proteins derived from IP experiments with 1- and 3-pHis antibodies, which indicates that neither of these analyses were able to access the entire pHis proteome alone. A combination of strategies will clearly be required to enable further exploration of this much-understudied area of the human phosphoproteome.

Nevertheless, the dataset generated here still contains a wealth of useful information. The ~300 sites of His phosphorylation have provided the first opportunity to examine amino acid motifs surrounding the site of modification. Consequently, two Lys-directed motifs have been revealed. The sequence of amino acids surrounding a phosphorylation site are crucial for kinase specificity. Hence, establishing such motifs can go some way to understanding the kinase-substrate interactions that modulate the network of phosphorylation in signal transduction pathways.

Within this dataset a total of 12 pHis peptides derived from kinases were identified. Three of these pHis peptides map to the identical region of the αC-β4 loop within the kinase catalytic domain, which is implicated in positioning of the αC-helix and thus catalytic activity. His phosphorylation of the HxN-His position could therefore act as a general regulatory mechanism of eukaryotic protein kinases. Further studies to establish the effects of this phosphorylation on kinase activity will however be required.

Given that His has a near neutral pKa and is thus partially protonated at physiological pH, the addition of a negatively charged phosphate group could result in a reversal of local net charge. This switch from positive to negative charge could conceivably cause a substantial conformational change within the protein, and will almost undoubtedly change the potential for local electrostatic interactions, thus regulating binding of the protein to other substrates. A positive to negative switch may also help regulate binding of proteins to nucleic acids, whereby phosphorylation of a His residue would inhibit binding of negatively charged nucleic acids. The identification of a number of RNA/DNA binding proteins in this

analysis, such as T-box transcription factor TBX4, RNA-binding protein 8A, Protein RRP5 homolog (also known as NFκB-binding protein) and Ribosomal L1 domain-containing protein 1, hints at this being a possible role for pHis in human cells. As a further consequence of the near neutral pKa of His, these residues can exhibit catalytically important transitions from a charged to uncharged side chain. Reversible His phosphorylation could therefore have a major impact on catalysis.

The identification of widespread His phosphorylation in human cells highlights the pressing need to identify the kinases and phosphatases involved in modulation of pHis in cellular signalling events. Biochemical evaluation and genomic annotation will be vital in this endeavour to refine the current understanding of eukaryotic signalling. Furthermore, identification of specific kinase-substrate interactions responsible for His phosphorylation may reveal druggable targets to modulate this activity. Given that pHis is known to be involved in cancer and tumour metastasis, identification of such targets may have therapeutic benefits, akin to the protein kinase inhibitors currently used in cancer treatments. As it is unlikely that pHis functions independently from pSer, pThr and pTyr, the effect of clinical inhibitors of protein kinases on global pHis in human cells, particularly those derived from disease models, may also provide an interesting avenue to establish the roles of pHis in human health and disease.

Chapter 6. Results IV: Analysis of the unexplored human phosphoproteome

6.1 Introduction

Besides the well-known phosphorylation of Ser, Thr and Tyr, and the previously discussed phosphorylation of acid-labile His, a number of other amino acid residues are known to be phosphorylated in biological systems. Phosphorylated Lys, Arg, Asp, Glu and Cys have all been reported, although very little is known about the extent of these phosphorylated residues in human systems, and what their roles might be[6-8]. The structures of these phosphorylated amino acids are shown in Figure 6.1.

Similarly to pHis, the phosphoramidate bond of pLys and pArg makes them acid-labile, which significantly hinders their analysis. Despite the fact that phosphorylation of lysine was first observed in rat liver cells 50 years ago[269], very little is known about this modification. There is evidence for *in vivo* phosphorylation of histone H1 on lysine[41], although the kinases responsible are unknown, with only partial purification of a possible candidate achieved[40]. There are also a few examples of pLys phosphatases isolated from mammalian systems, and the phosphatase PHPT1 has also been shown to dephosphorylate pLys *in vitro*[270-272]. Much of the current research regarding pArg is focused on bacteria, with the first protein arginine kinase identified less than 10 years ago[273], and recent roles being discovered for pArg in bacterial stress response[234]. However, there are also several older reports of pArg kinase and phosphatase activity in vertebrates[274-278]. In general, reported pArg kinases are isolated from nuclear tissue, whilst the few known substrates tend to be basic proteins such as histones and myelin.

The phosphorylation of Asp and Glu to form phosphoanhydrides increases their susceptibility to hydrolysis, and as such they are reported to be very unstable, at least in their free forms[7]. The role of pAsp in the two-component signalling system is well established; pAsp relays a signal from a sensor histidine kinase to a response regulator, which modulates a cellular response, for example transcription[16]. The two-component signalling system is well known in microorganisms and plants, although an analogous system is not thought to function in mammals. However, in human cells pAsp is known to occur as an enzymatic intermediate in a number of enzyme catalysed reactions[279, 280], as

well as pAsp proteins having previously been identified in the study of human prostate cancer cell lines[218]. According to the literature, pGlu is best known for its free form gamma glutamyl phosphate, which is an important intermediate in the synthesis of the amino acids glutamine and proline[281, 282], as well as being involved in the reaction to form glutathione[283]. In terms of its currently known roles in protein phosphorylation, pGlu is found in the nuclear protein prothymosin, where it is reportedly involved in production, processing or export of RNA[284, 285], and has also been reported to occur in α2 chains of collagen[286].

It was hypothesised that given the UPAX method had proven to be successful for pHis analysis, it may also have preserved other acid-labile phosphate groups, which could be identified through additional interrogation of the data. This chapter describes the identification and bioinformatics characterisation of non-canonical sites of phosphorylation from the data acquired following treatment of HeLa cells with PHPT1 siRNA and NT siRNA with subsequent SAX fractionation and analysis by LC-MS/MS (as detailed in *Chapter 5. Results III*).

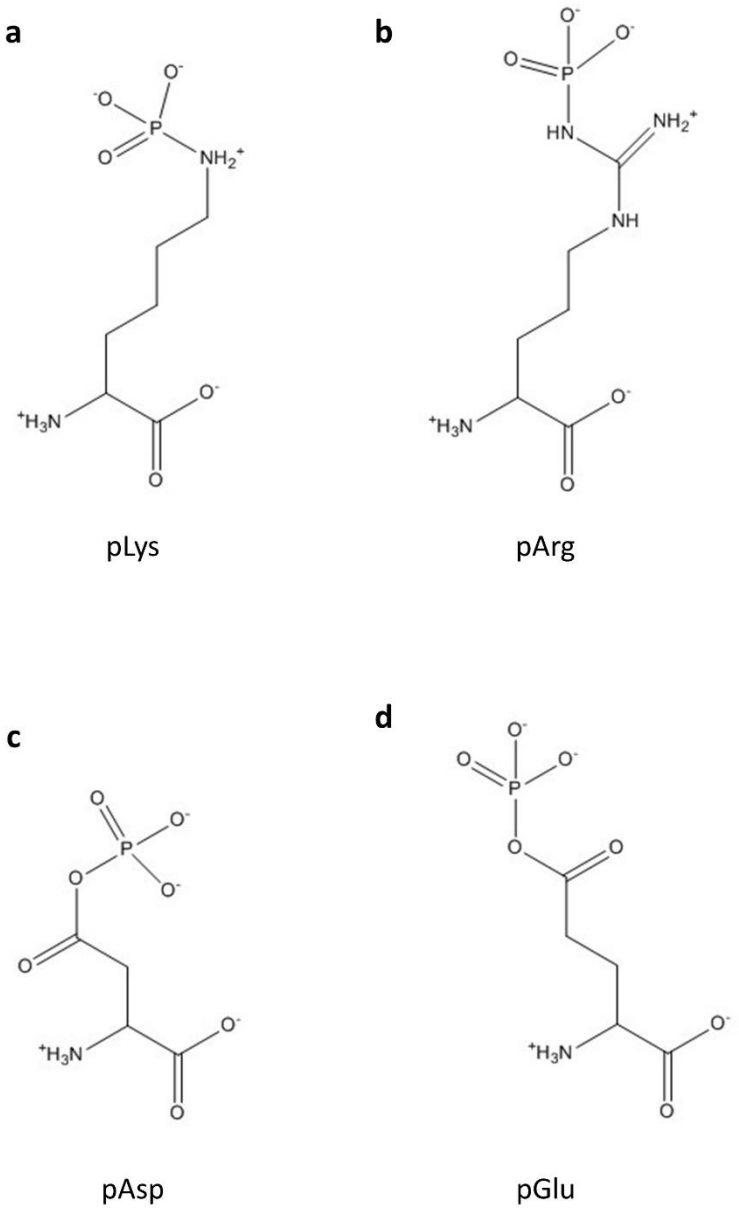


Figure 6.1 Phosphorylated amino acid structures a) phospholysine **b)** phosphoarginine **c)** phosphoaspartate and **d)** phosphoglutamate

6.2 Results and Discussion

6.2.1 Identification of non-canonical sites of phosphorylation

To perform a MASCOT search with all possible sites of phosphorylation considered as variable modifications would result in a massively increased theoretical search space, making the search very slow and causing a huge increase in false discovery rates, and thus decreasing the number of (phospho)peptides identified. Therefore, it was decided that a better approach was to perform searches with phosphorylated Ser, Thr, Tyr and one additional site of modification at a time. The inclusion of Ser, Thr and Tyr phosphorylation was important as a large number of these sites had previously been identified in the MASCOT searches for pHis, with the additional observation that numerous pHis-containing peptides were co-phosphorylated on Ser, Thr or Tyr.

MASCOT searches were performed in the same way as they had been for pHis but with the other phosphosites. An initial evaluation of the extent of non-canonical phosphorylation was carried out by performing MASCOT searches for each site of phosphorylation (pLys, pArg, pAsp, pGlu and pCys) for one set of SAX fractions. There was no evidence for pCys in these samples, but the other four sites of phosphorylation warranted further investigation. For pLys, pArg, pAsp and pGlu MASCOT searches were subsequently carried out for all three bioreplicates across the two conditions (PHPT1 siRNA and NT siRNA). Phosphosite localisation confidence was assessed using the ptmRS node in Proteome Discoverer, with a 1% FLR cut-off used as previously described. The number of non-canonical phosphosites identified in each of these searches is shown in Table 6.1, with the number of pSer, pThr and pTyr sites (identified in the search including pHis) shown for comparison.

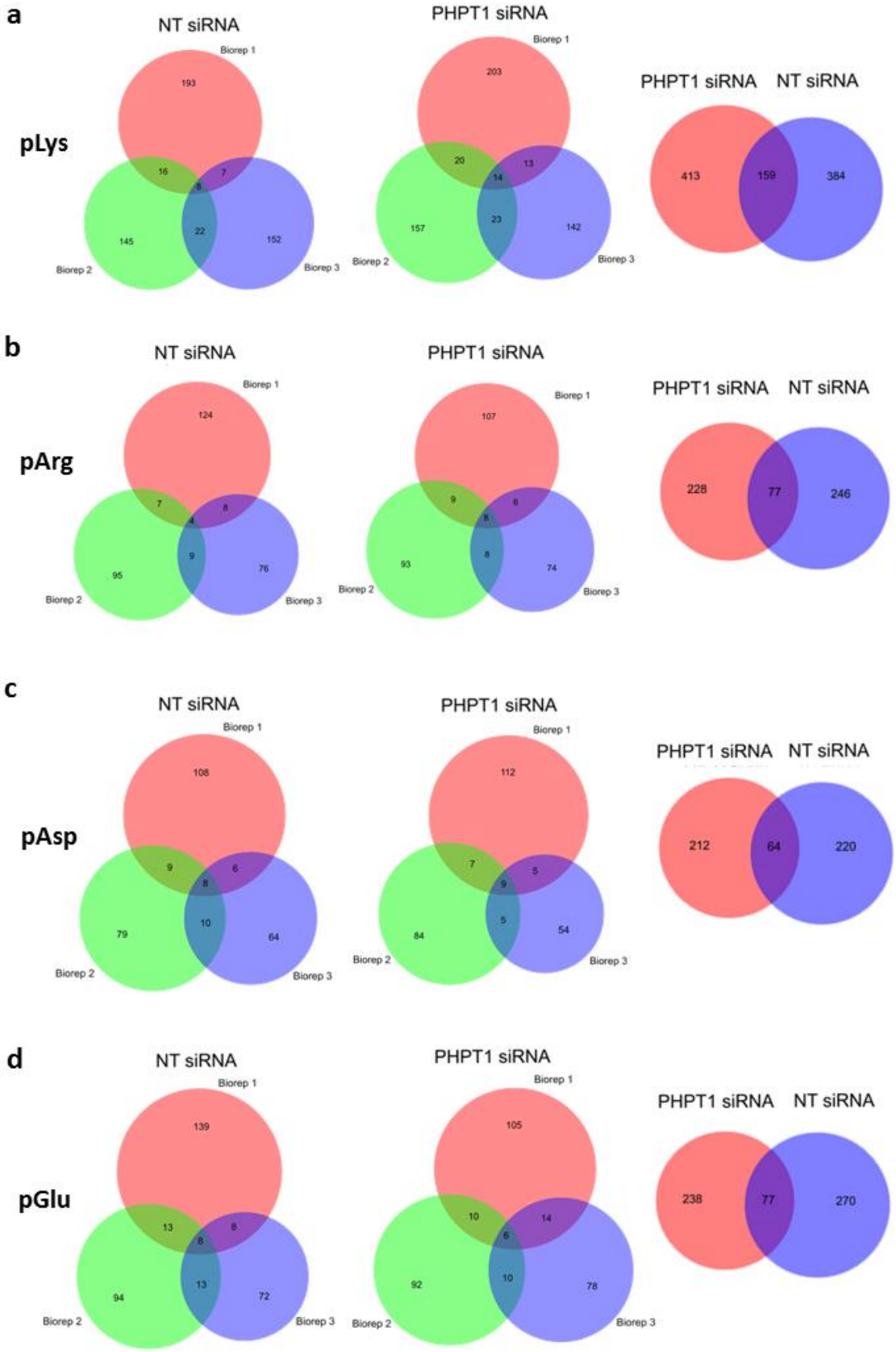
A surprisingly large number of pLys, pArg, pAsp and pGlu sites are identified in these searches, with the number of all of these sites exceeding the number of pHis and pTyr sites identified. Interestingly, for all sites except pLys there are a greater number of phosphosites identified in the NT siRNA condition compared to PHPT1 siRNA. The increase in pLys sites upon treatment with PHPT1 siRNA aligns with the previously published observation of dephosphorylation of pLys by PHPT1 *in vitro*[272], suggesting that PHPT1 phosphatase activity towards pLys may also be relevant *in vivo*, and perhaps PHPT1 regulation of pLys is more prevalent than regulation of pHis.

Table 6.1 Number of phospho-sites, peptides and proteins identified following MASCOT searches for modification of acid-labile residues. MASCOT searches of all 16 SAX fractions for each of the three bioreplicates in both conditions (PHPT1 siRNA and NT siRNA) were conducted using Proteome Discoverer with 5% FDR. Searches for pHis, pArg, pLys, pAsp and pGlu were conducted separately with pSer, pThr and pTyr additionally included as variable modifications for each pX. Numbers for pSer, pThr and pTyr are taken from the search with pHis. All sites meet the 1% FLR cut-off based on ptmRS score.

Phosphosite	Condition (siRNA)	Unique pX peptides	Unique pX sites	Unique pX proteins
pHis	**PHPT1**	164	172	164
	NT	181	186	176
pArg	**PHPT1**	305	324	329
	NT	323	345	333
pLys	**PHPT1**	572	618	555
	NT	543	592	533
pAsp	**PHPT1**	276	287	264
	NT	284	294	274
pGlu	**PHPT1**	315	334	304
	NT	347	374	324
pSer	**PHPT1**	2655	2889	1575
	NT	3024	3321	1745
pThr	**PHPT1**	626	681	560
	NT	666	707	573
pTyr	**PHPT1**	152	159	148
	NT	169	175	165

The degree of overlap for phosphopeptide IDs between the PHPT1 siRNA and NT siRNA conditions is similar for all non-canonical sites (Figure 6.2 a-d), ranging between 13 and 17%. For pSer, pThr and pTyr phosphopeptides (identified in the search including pHis) overlap between the two conditions is 42%, 22% and 13% respectively. The greater degree of overlap between the pSer and pThr peptides is perhaps unsurprising given that many pSer/pThr mediated signalling events would not be expected to change between the two treatment conditions. The overlap in phosphopeptides IDs between bioreplicates for non-canonical sites (pLys, pArg, pAsp and pGlu) and pTyr is fairly low, with the majority of phosphopeptides identified in only one bioreplicate. This is probably due to the stochastic nature of MS analysis, with low abundance phosphopeptides less likely to be identified in all samples, i.e. these phosphopeptides may not be selected for MS2 in some samples if

they are below the required intensity threshold. The highest confidence identifications will be those observed in more than sample, but for the purposes of this work all identifications were kept for further bioinformatics analysis. It would be wise to assess the number of times a given pX peptide of interest has been observed in these experiments as part of the criteria for selecting pX proteins that warrant further biochemical investigation. The results of each of the searches are discussed in more detail in the subsequence sections of this chapter. The full list of peptide and protein IDs for each site of phosphorylation can be found in the Appendix, along with example tandem MS spectra for peptides bearing each type of phosphorylation.


a
NT siRNA
Biorep 1
193
16
7
8
145
22
152
Biorep 2
Biorep 3
pLys
PHPT1 siRNA
Biorep 1
203
20
13
14
157
23
142
Biorep 2
Biorep 3
PHPT1 siRNA
NT siRNA
413
159
384
b
NT siRNA
Biorep 1
124
7
8
4
95
9
76
Biorep 2
Biorep 3
pArg
PHPT1 siRNA
Biorep 1
107
9
6
8
93
8
74
Biorep 2
Biorep 3
PHPT1 siRNA
NT siRNA
228
77
246
c
NT siRNA
Biorep 1
108
9
6
8
79
10
64
Biorep 2
Biorep 3
pAsp
PHPT1 siRNA
Biorep 1
112
7
5
9
84
5
54
Biorep 2
Biorep 3
PHPT1 siRNA
NT siRNA
212
64
220
d
NT siRNA
Biorep 1
139
13
8
8
94
13
72
Biorep 2
Biorep 3
pGlu
PHPT1 siRNA
Biorep 1
105
10
14
6
92
10
78
Biorep 2
Biorep 3
PHPT1 siRNA
NT siRNA
238
77
270

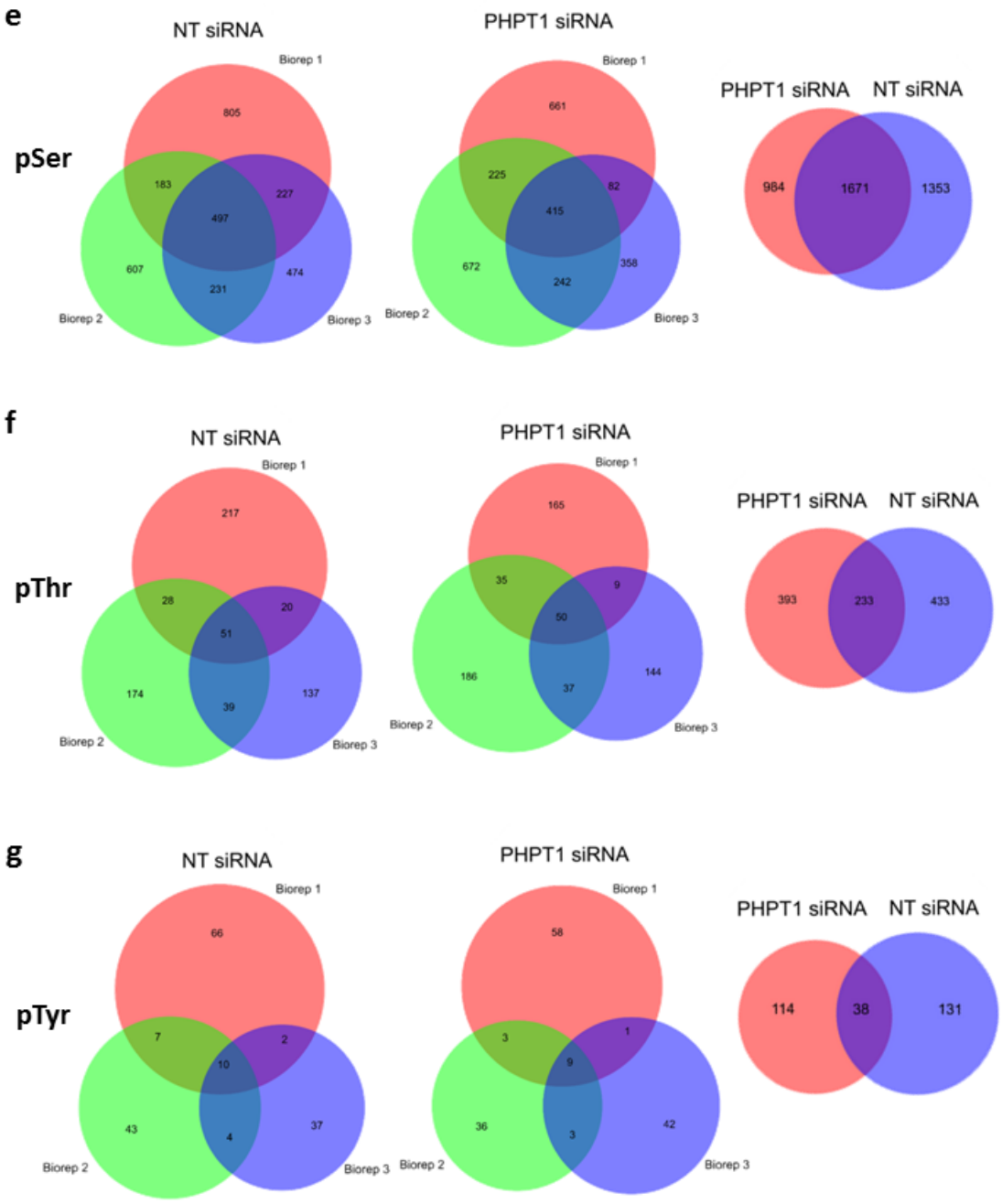


Figure 6.2 Overlap in phosphopeptide IDs between bioreplicates for each condition (PHPT1 siRNA and NT siRNA) and between conditions, shown for each site of phosphorylation. a) pLys, **b)** pArg, **c)** pAsp, **d)** pGlu, **e)** pSer, **f)** pThr and **g)** pTyr. Numbers for pSer, pThr and pTyr are taken from the search with pHis. Overlap for pHis peptides shown in *Chapter 5. Results III: Figure 5.15 b.*

6.2.2 Phospholysine

Contrary to the trend observed for all other phosphosites, a greater number of pLys peptides are observed from proteins extracted from the PHPT1 siRNA treated samples (572 peptides, 618 sites) compared to the NT siRNA treated samples (543 peptides, 592 sites). This corresponds to 555 and 533 pLys proteins identified in the PHPT1 siRNA and NT siRNA conditions respectively. Of the non-canonical phosphosites searched for in the data, pLys peptides are the most prevalent, with over 3 times as many pLys peptides compared to pHis. Of the 572 peptides identified in the PHPT1 siRNA condition 28% (159 peptides) are also identified in the NT siRNA condition (Figure 6.2 a).

The increase in the number of pLys peptides (and proteins) identified following PHPT1 siRNA treatment suggest that PHPT1 is acting as a pLys phosphatase. This aligns with previous evidence for dephosphorylation of pLys by PHPT1 *in vitro*[272], and thus suggests that the phosphatase activity of PHPT1 towards pLys is also relevant *in vivo*. Histone H1 has previously been reported to be phosphorylated on Lys, by as yet unknown kinases[40, 41]. Interestingly, there are 3 pLys peptides identified in this study that belong to histones (H1x, H2A.J and H2B).

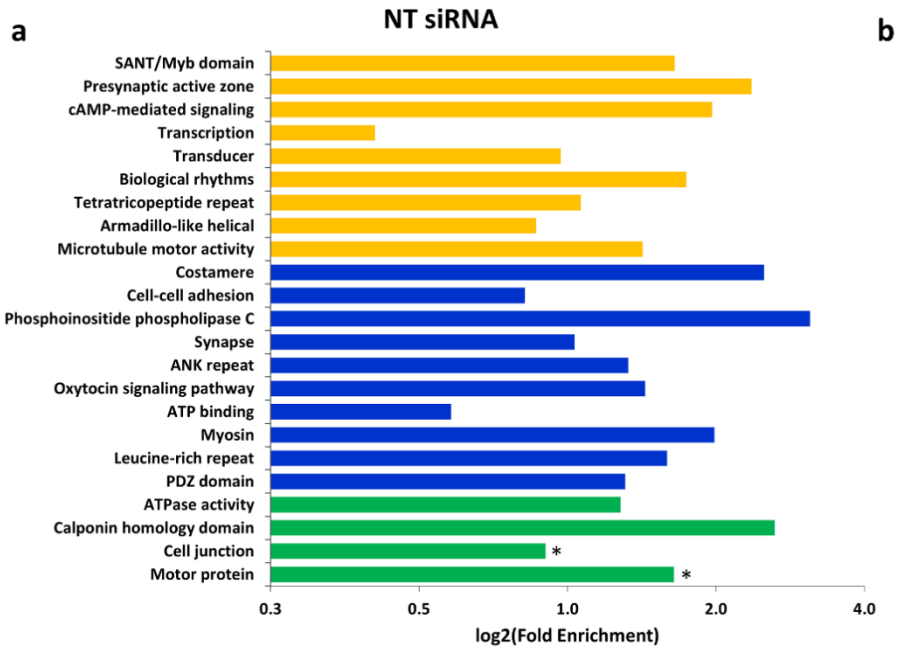
Functional annotation clustering was performed using DAVID to assess enrichment of pLys proteins identified in each of the two conditions compared to the background of all proteins identified in these studies (Figure 6.3). Significant enrichment for pLys was observed in motor protein (18 proteins, 3.1 fold, P=5.21E-5), cell junction (35 proteins, 1.9 fold, P=5.13E-4), calponin homology domain (7 proteins, 6.2 fold, P=6.52E-4) and ATPase activity (19 proteins, 2.4 fold, P=7.78E-4). PDZ domain (15 proteins, 2.5 fold, P=2.76E-3), leucine-rich repeat (11 proteins, 3 fold, P=3.17E-3) and myosin (8 proteins, 3.9 fold, P=3.47E-3) are also enriched, with a small but significant enrichment also observed in ATP binding (51 proteins, 1.5 fold, P=3.6E-3).

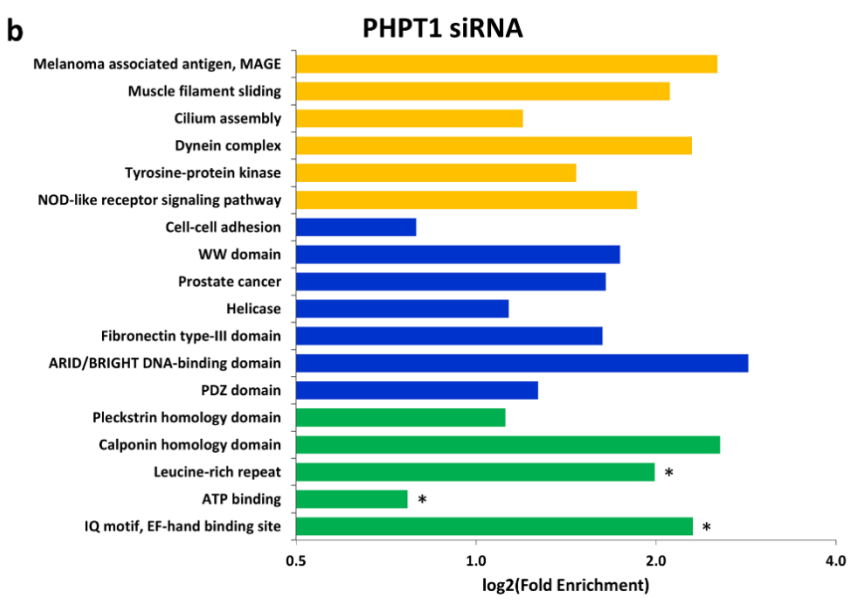
Following treatment with PHPT1 siRNA IQ motif with EF-hand binding site shows significant enrichment (15 proteins, 4.9 fold, P=1.10E-6), which was not observed in the NT siRNA condition. Additionally, there is increased significance associated with ATP binding (69 proteins, 1.7 fold, P=1.22E-5) and leucine-rich repeat (12 proteins, 4 fold, P=1.63E-4) compared to the NT siRNA condition. Calponin homology domain (7 proteins, 5.9 fold,

P=8.44 E-4), pleckstrin homology domain (23 proteins, 2.2 fold, P=8.61E-4) and PDZ domain (15 proteins, 2.4 fold, P=3.53E-3) are also enriched.

There were eight KEGG pathways significantly enriched under NT siRNA conditions (P<5E-2). Most notable were tight junction (12 proteins including myosins, P=1.73E-3) and oxytocin signalling pathway (11 proteins, P=6.52E-3). The tight junction is one of ten KEGG pathways enriched following PHPT1 knockdown (13 proteins, P=1.04E-3), although the oxytocin signalling pathway enriched under control conditions was not observed following PHPT1 depletion. However, there was a 3-fold enrichment of the prostate cancer pathway (P=6.92E-3), represented by 9 proteins.

All sequences identified across the two conditions were assessed using the Motif-X software (Figure 6.4). There appears to be a strong preference for the hydrophobic residues Leu or Ile at the -2 or -7 positions with respect to the phosphorylated Lys residue; the top scoring motifs have Leu in the -2 position (score 14.45, fold enrichment 1.80) and Leu in the -7 position (score 10.84, fold enrichment 1.81). The positions surrounding the phosphorylated Lys seem to favour hydrophobic residues, particularly Leu and Val, along with Glu. Leu directed motifs also emerge for peptides identified in the first eight fractions, with Leu in the -2 position having a particularly high score (score 15.95, fold enrichment 2.17). For sequences identified in the last eight fractions the motif is Glu directed (Glu -1, score 7.81, fold enrichment 1.92), reflecting a possible bias due to the charge-based separation which result in later fractions containing peptides with more acidic residues.





KEGG Pathway	Proteins	P Value	Fold Enrichment
Tight junction	12	1.73E-3	3.01
Oxytocin signalling pathway	11	6.52E-3	2.71
Dilated cardiomyopathy	7	1.29E-2	3.52
Platelet activation	9	2.66E-2	2.48
Thyroid hormone synthesis	6	3.25E-2	3.30
Inflammatory mediator regulation of TRP channels	7	3.67E-2	2.79
Hypertrophic cardiomyopathy (HCM)	6	3.87E-2	3.15
Vascular smooth muscle contraction	7	4.85E-2	2.61

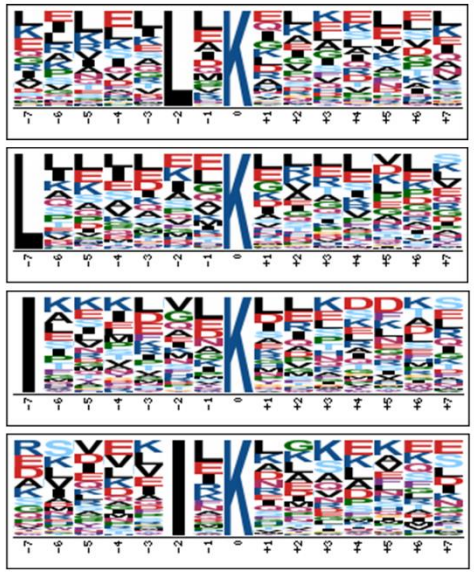
KEGG Pathway	Proteins	P Value	Fold Enrichment
Tight junction	13	1.04E-3	3.00
Prostate cancer	9	6.92E-3	3.13
NOD-like receptor signalling pathway	7	1.11E-2	3.63
Toll-like receptor signalling pathway	8	1.91E-2	2.88
Insulin signalling pathway	11	2.25E-2	2.25
Adipocytokine signalling pathway	7	2.32E-2	3.10
Chagas disease (American trypanosomiasis)	8	3.56E-2	2.54
Herpes simplex infection	12	3.76E-2	1.98
PI3K-Akt signalling pathway	16	4.08E-2	1.73
Hepatitis B	10	4.53E-2	2.10

Figure 6.3 Enrichment analysis of pLys proteins performed using DAVID a) proteins identified in NT siRNA condition, with enriched KEGG pathways shown in the table **b)** proteins identified in PHPT1 siRNA condition, with enriched KEGG pathways shown in the table. P value indicated by colour: yellow P<0.05, blue P<0.01, green P<0.001 with asterisks defining clusters where the Benjamini adjusted P-value <0.1

a

All fractions

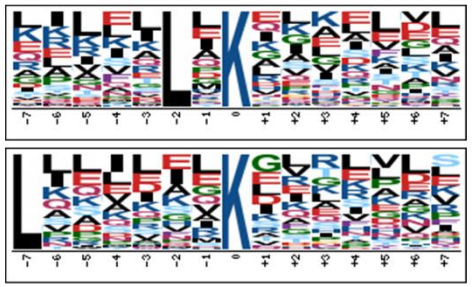
Motif	Score	Matches in dataset	Fold increase
.....L.K.....	14.45	187	1.80
L.....K.....	10.84	137	1.81
I.....K.....	7.39	69	2.02
.....I.K.....	6.41	65	1.94



b

Fractions 1-8

Motif	Score	Matches in dataset	Fold increase
.....L.K.....	15.95	125	2.17
L.....K.....	8.96	81	2.02



c

Fractions 9-16

Motif	Score	Matches in dataset	Fold increase
.....EK.....	7.81	80	1.92

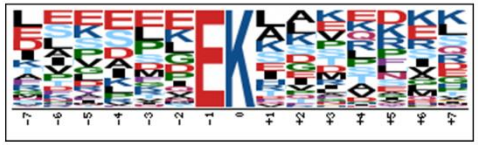


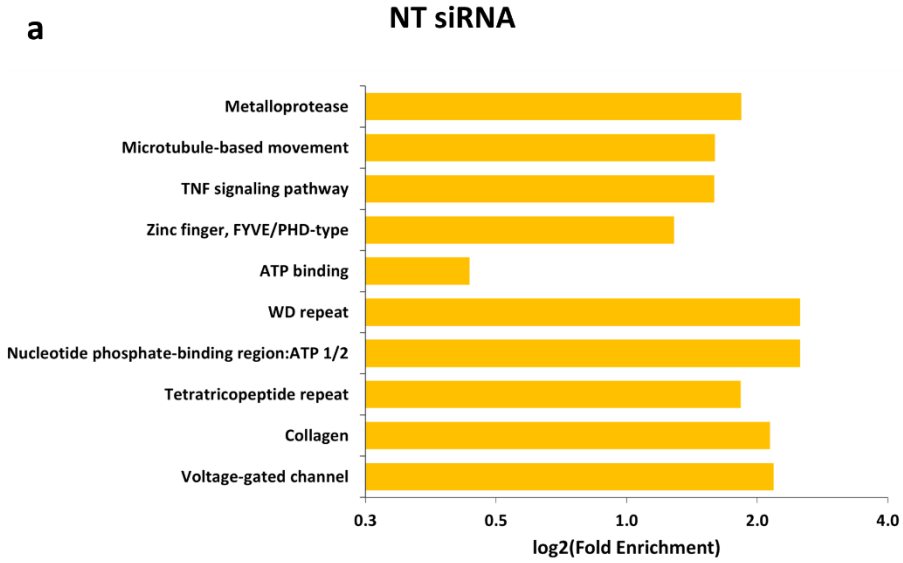
Figure 6.4 Extracted motifs from pLys peptide data a) Motifs from pLys peptides in all 16 SAX fractions **b)** Motifs from pLys peptides in SAX fractions 1-8 **c)** Motif from pLys peptides in SAX fractions 9-16. A higher score indicates a more significant/more specific motif. Fold increase compared to background of all sequences in the human proteome. Significance set at 1E-5 (not corrected for multiple hypotheses, corresponds to P value of 3E-4 with Bonferroni correction).

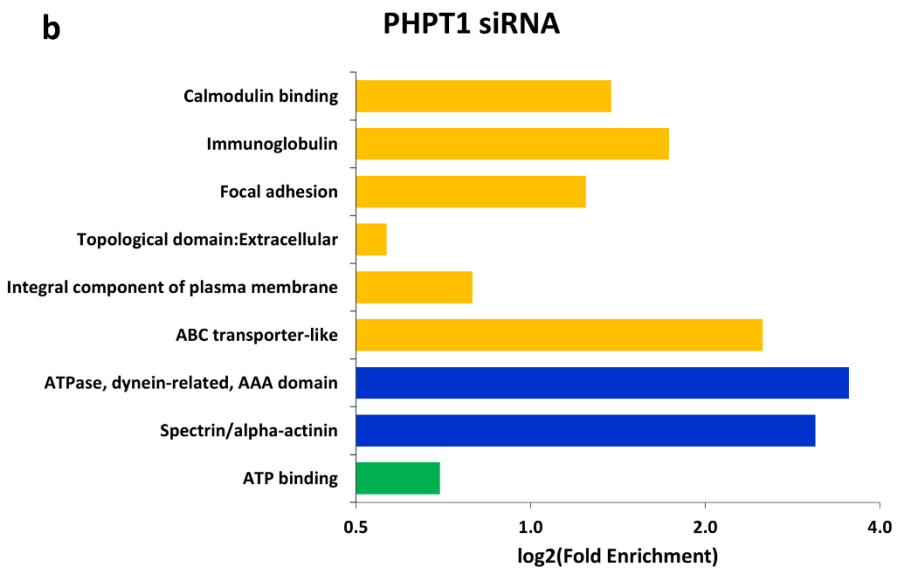
6.2.3 Phosphoarginine

A total of 551 pArg containing peptides were identified across both conditions, corresponding to 585 pArg proteins. Of the 323 pArg peptides identified in the NT siRNA condition 24% (77 peptides) were also identified in the PHPT1 siRNA condition. The very few examples of pArg on protein substrates in mammalian systems occur on histones or myelin; in this analysis one pArg-containing peptide derived from Histone H2A was identified.

Functional annotation clustering (Figure 6.5) of the proteins identified in the NT siRNA condition revealed only clusters in the lowest category of significance (5E-2>P>1E-2). These clusters included voltage gated channel (5 proteins, P=2.33E-2), collagen (5 proteins, P=2.57E-2) and tetratricopeptide repeat (6 proteins, P=2.58E-2). However, following treatment with PHPT1 siRNA a number of significant enrichments are observed. Most notable is ATP binding (48 proteins, 1.6 fold, P=6.99E4), with spectrin/alpha-actin (5 proteins, 8.5 fold, P=2.42E-3) and ATPase, dynein related, AAA domain (4 proteins, 11.6 fold, P=4.29E-3) also enriched, although the latter two are based on a relatively small number of proteins.

In both conditions three KEGG pathways were enriched (P<5E-2), although they were different pathways for each condition. The mTor signalling pathway (5 proteins, 4.14 fold, P=3.06E-2) and PI3K-Akt signalling pathway (11 proteins, 2.08 fold, P=3.52E-2) are enriched in the NT siRNA conditions, whilst focal adhesion (9 proteins, 2.37 fold, P=3.32E-2) is the most significantly enriched pathway in the PHPT1 siRNA condition.





KEGG Pathway	Proteins	P Value	Fold Enrichment
mTOR signalling pathway	5	3.06E-2	4.14
PI3K-Akt signalling pathway	11	3.52E-2	2.08
TNF signalling pathway	6	4.58E-2	3.02

KEGG Pathway	Proteins	P Value	Fold Enrichment
Focal adhesion	9	3.32E-2	2.37
Ras signalling pathway	8	3.40E-2	2.56
ABC transporters	4	3.67E-2	5.34

Figure 6.5 Enrichment analysis of pArg proteins performed using DAVID a) proteins identified in NT siRNA condition, with enriched KEGG pathways shown in the table **b)** proteins identified in PHPT1 siRNA condition, with enriched KEGG pathways shown in the table. P value indicated by colour: yellow P<0.05, blue P<0.01, green P<0.001.

Motif-X analysis (Figure 6.6) of all sequences reveals two motifs: Glu in the +1 position (score 6.49, fold increase 1.85) and Leu in the -2 position (score 6.70, fold enrichment 1.73). There also appears to be a general preference for Glu, Leu, Ile, and Ala in the positions surrounding the phosphorylated residue. When sequences from the first or last eight fractions are analysed separately no motifs are found, indicating that neither of these identified motifs are being biased as a result of the charge-based SAX fractionation strategy.

All fractions

Motif	Score	Matches in dataset	Fold increase
.....RE.....	6.49	74	1.85
.....L.R.....	6.70	91	1.73

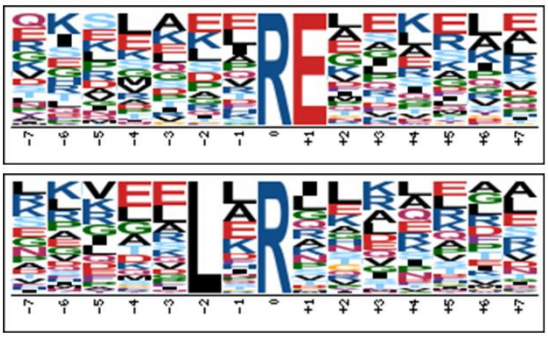


Figure 6.6 Extracted motifs from pArg peptide data for peptides identified across all SAX fractions. A higher score indicates a more significant/more specific motif. Fold increase compared to background of all sequences in the human proteome. Significance set at 1E-5 (not corrected for multiple hypotheses, corresponds to P value of 3E-4 with Bonferroni correction).

6.2.4 Phosphoaspartate

In this analysis, a total of 496 pAsp peptides (corresponding to 474 proteins) were identified. This is over 1.5 times the number of pHis-peptides/proteins identified, which aligns with previous LC-MS/MS analysis of pHis/pAsp in a human prostate cancer cell line where a greater number of pAsp peptides were detected compared to pHis (75 pAsp proteins compared to 20 pHis proteins[218]). Of these previously identified pAsp proteins, seven were identified in this experiment (Table 6.2). There is however, disagreement in all cases over the phosphorylated Asp residue within these proteins, with different pAsp peptides identified in this experiment compared to the Lapek *et al.* data for each of the proteins. Of all of the pAsp peptides identified by Lapek *et al*. only one is identified in this data, with the phosphorylation site localised to a Ser residue in this analysis: SLSSpSLDDTEVKK from VAMP-associated protein with a ptmRS score of 76.69 for residue S5. The discrepancies between these two data sets highlights the difficulty in obtaining comprehensive and confident coverage of this unexplored area of the phosphoproteome.

Table 6.2 Previously identified pAsp proteins. Proteins previously identified by Lapek *et al.* in human prostate cancer cell lines that are also observed in this analysis with at least one pAsp site. The pAsp sites observed in this analysis are not the same as those previously identified.

Protein Description	Accession	pAsp residue	Condition (siRNA)
26S proteasome non-ATPase regulatory subunit 2	Q13200	784	PHPT1
Calumenin	O43852	41; 250; 256	Both
Eukaryotic translation initiation factor 2	P20042	4	NT
Exportin 1	O14980	681	PHPT1
Nesprin 1	Q8NF91	8307	PHPT1
Neuroblast differentiation-associated protein AHNAK	Q09666	4432; 4563; 4990; 5742	Both
Stress induced phosphoprotein 1	P31948	70	Both

Interestingly, a peptide from PHPT1 was identified in three out of the six bioreplicates as being phosphorylated on an Asp residue. The same peptide is also observed in the remaining three bioreplicates but the ptmRS scores are below the 1% FLR cut-off (although the Asp site is still the most likely position of phosphorylation in all cases). Figure 6.7 shows a representative MS/MS spectrum for the pAsp peptide (AVApDLALIPDVDIDSDGVFK, *m/z* 718.35), with the pAsp site localised to position 4. This pAsp site corresponds to residue D5 in the protein sequence, which has not previously been identified as a modified site.

Analysis with DAVID (Figure 6.8) for proteins identified in the NT siRNA condition revealed significant enrichment for tetratricopeptide repeat (12 proteins, 7.8 fold, P=2.93E-7) and cilium (11 proteins, 3.3 fold P=1.88E-3), with enrichment also observed for cell junction (20 proteins, 2.1 fold, P=2.63E-3), cell cycle arrest (9 proteins, 3.7 fold, P=2.79E-3) and microtubule (14 proteins, 2.3 fold, P=7.09E-3). Proteins identified in the PHPT1 siRNA condition also show significant enrichment for tetratricopeptide repeat (7 protein, 9.6 fold, P=6.74E-5). Treatment with PHPT1 siRNA also results in significant enrichment of ATP binding (36 proteins, 2.1 fold, P=2.40E-5) and ATPase activity (13 proteins, 3.3 fold, P=5.53E-4), with motor protein (9 proteins, 3.2 fold P=6.60E-3), helicase (10 proteins, 2.9 fold, P=6.73E-3) and disulphide bond (32 proteins, 1.6 fold, P=6.78E-3) also enriched. There were no significantly enriched KEGG pathways identified for proteins in either condition

Motif-X analysis using all sequences did not reveal any motifs, however, when only the sequences identified in fractions 1-8 were used, one motif does emerge (Figure 6.8 c). The identified motif has Lys in the +4 position (score 6.64, fold increase 2.85). Additionally, there appears to be a preference for Lys and hydrophobic residues Leu, Ile and Ala in the positions surrounding the phosphorylated Asp residue. Motif analysis of the small number of pAsp peptides previously identified also highlighted a preference for Asp, Lys and hydrophobic residues in close proximity to pAsp[218].

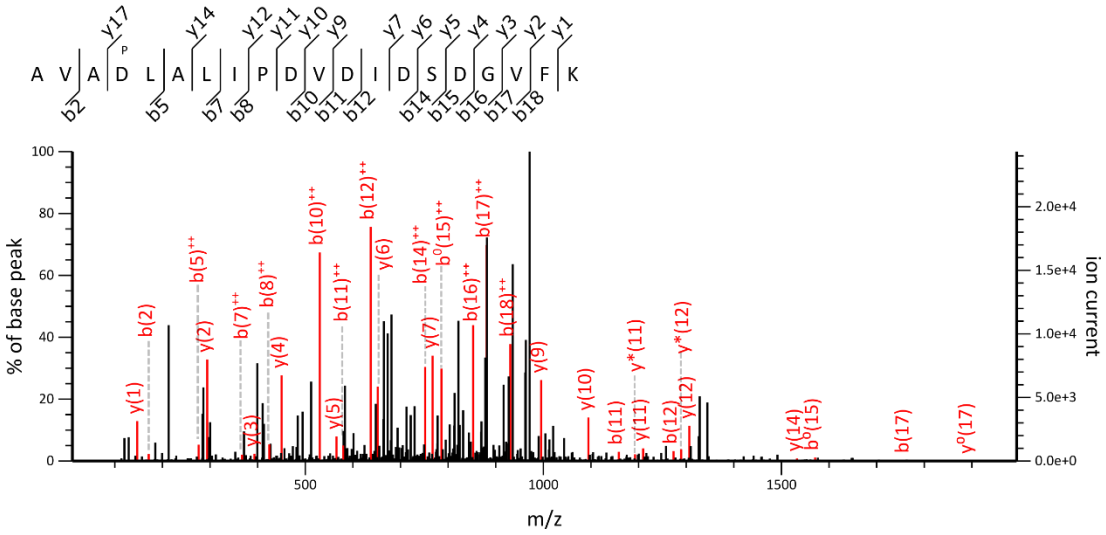


Figure 6.7 Tandem mass spectrum of triply charged pAsp peptide from PHPT1. Spectrum generated by HCD of peptide ion at m/z 718.35 (AVApDLALIPDVDIDSDGVFK); LC-MS/MS analysis with Orbitrap Fusion of SAX fraction 8 of NT siRNA treated sample. MASCOT ion score: 32, ptmRS score for residue D4: 99.94.

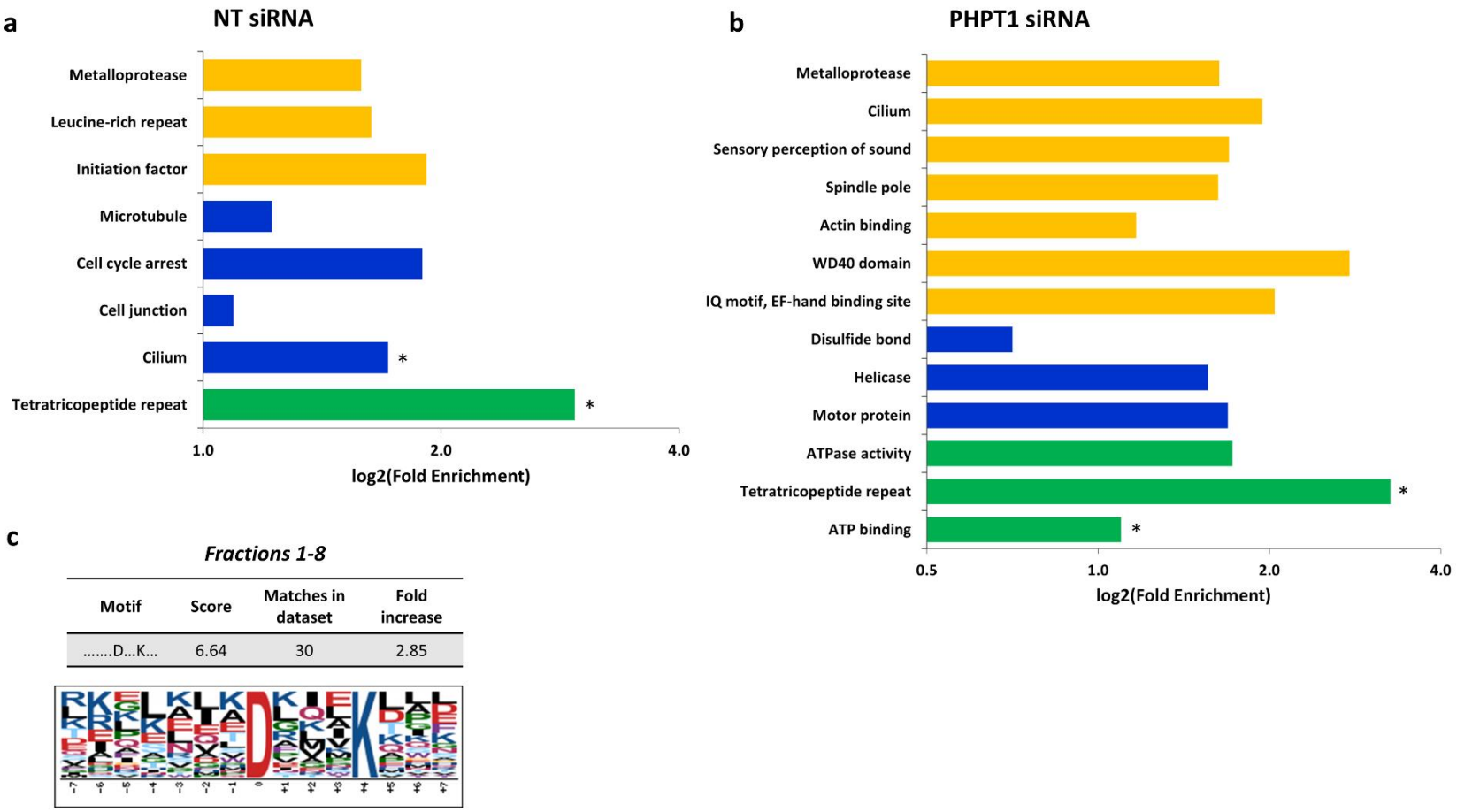


Figure 6.8 Enrichment analysis of pAsp proteins performed with DAVID and a pAsp motif a) Enrichment analysis for proteins identified in NT siRNA condition **b)** Enrichment analysis for proteins identified in PHPT1 siRNA condition. P value indicated by colour: yellow P<0.05, blue P<0.01, green P<0.001 with asterisks defining clusters where the Benjamini adjusted P-value <0.1 **c)** Motif analysis for sequences in fractions 1-8 (no motifs found when all sequences searched). Significance set at 1E-5 (not corrected for multiple hypotheses, corresponds to P value of 3E-4 with Bonferroni correction).

6.2.5 Phosphoglutamate

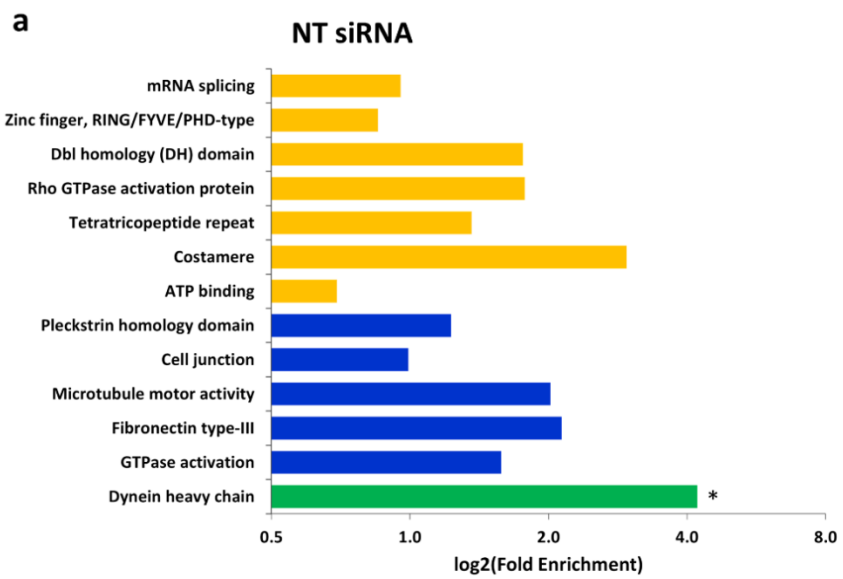
A total of 585 pGlu-containing peptides were identified across the two conditions, corresponding to 551 pGlu-containing proteins. Of the 347 pGlu peptides identified in the NT siRNA condition 22% (77 peptides) were also identified in the PHPT1 condition, which is a similar degree of overlap as is observed for the other non-canonical phosphorylation sites. One of the few currently known examples of pGlu protein phosphorylation occurs on α2 chains of collagen[286]. In this data set one pGlu peptide is observed from α2 collagen, with a further four pGlu peptides from other α1 and α4 collagen chains.

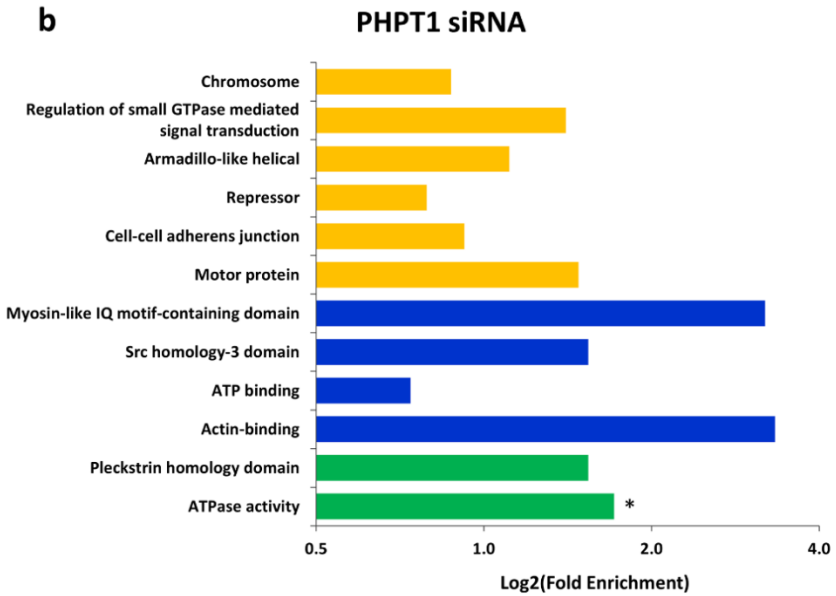
Functional annotation clustering was performed using DAVID (Figure 6.9). Proteins identified in the NT siRNA condition are significantly enriched for dynein heavy chain (5 proteins, 18.5 fold, P=8.79E-5). Also enriched are GTPase activation (13 proteins, 3 fold, P=1.34E-3), fibronectin type-III (8 proteins, 4.4 fold, P=2.11E-3) and microtubule motor activity (8 proteins, 4.1 fold, P=3.35E-3). Following treatment with PHPT1 siRNA, ATPase activity is significantly enriched (15 proteins, 3.3 fold, P=1.84E-4). Pleckstrin homology domain exhibits a similar degree of enrichment in both the PHPT1 siRNA and NT siRNA samples (PHPT1 siRNA: 15 proteins, 2.9 fold, P=3.38E-4; NT siRNA: 14 proteins, 2.3 fold, P=6.79E-3) as does ATP binding (PHPT1 siRNA: 32 proteins, 1.7 fold, P=5.10E-3; NT siRNA: 32 proteins, 1.6 fold, P=8.03E-3). Proteins in the PHPT1 siRNA condition are also enriched for actin-binding (5 proteins, 10 fold, P=1.23E-3), Src homology-3 domain (10 proteins, 2.9 fold, P=7.32E-3) and Myosin-like IQ motif-containing domain (4 proteins, 9.2 fold, P=8.47E-3).

There are no KEGG pathways enriched in the NT siRNA condition, however pGlu proteins in the PHPT1 siRNA sample map to a number of enriched KEGG pathways. There are 27 significantly enriched KEGG pathways with P<0.01, with a further 20 with P<0.05. The pathways with highest significance and represented by more than eight proteins are shown in Figure 6.9 b. Most notable are platelet activation (11 proteins, 4.7 fold, P=7.81E-3), insulin signalling pathway (11 proteins, 3.8 fold, P=4.67E-3) and AMPK signalling pathway (10 proteins, 4 fold, P=7.38 E-4).

Motif-X analysis (Figure 6.10) of all sequences across the two conditions reveals a strong preference for Lys in the +1 position (score 12.78, fold increase 2.22), with a second motif where the preference is for Lys in the -3 position (score 6.09, fold increase 1.91).

Performing the analysis with only peptides in the first eight SAX fractions also reveals the motif with Lys in the +1 position (score 10.63, fold increase 2.73), plus a second different motif with Lys in the +2 position (score 6.17, fold increase 2.61). There is also a general trend towards Leu and Ile in the positions surrounding the pGlu residue.





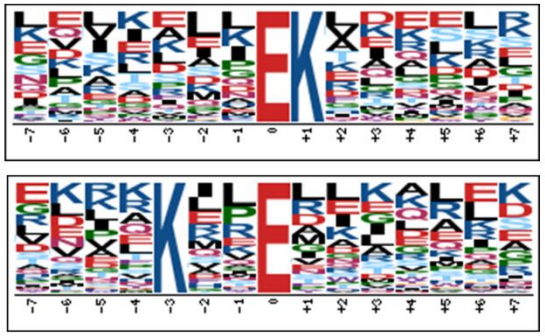
KEGG Pathway	Proteins	P Value	Fold Enrichment
Amoebiasis	10	5.44E-5	5.57
Platelet activation	11	7.81E-5	4.74
Insulin signalling pathway	11	4.67E-4	3.83
AMPK signalling pathway	10	7.38E-4	3.98
cGMP-PKG signalling pathway	10	1.56E-3	3.59
Influenza A	10	3.02E-3	3.26
Epstein-Barr virus infection	12	3.55E-3	2.75
PI3K-Akt signalling	13	6.92E-3	2.39

Figure 6.9 Enrichment analysis of pGlu proteins performed with DAVID a) proteins identified in NT siRNA condition **b)** proteins identified in PHPT1 siRNA condition, with enriched KEGG pathways shown in the table. P value indicated by colour: yellow P<0.05, blue P<0.01, green P<0.001 with asterisks defining clusters where the Benjamini adjusted P-value <0.1

a

All Fractions

Motif	Score	Matches in dataset	Fold increase
.....EK.....	12.78	97	2.22
....K..E.....	6.09	63	1.91



b

Fractions 1-8

Motif	Score	Matches in dataset	Fold increase
.....EK.....	10.63	53	2.73
.....E.K.....	6.17	32	2.61

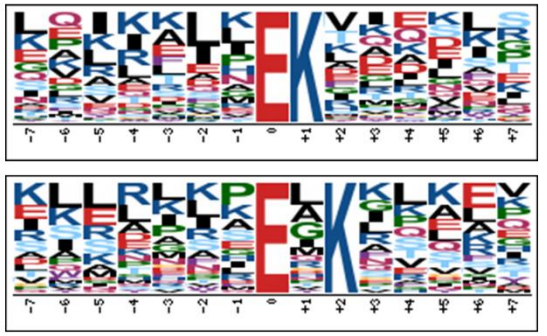


Figure 6.10 Extracted motifs from pGlu peptide data a) Motifs from pGlu peptides in all 16 SAX fractions **b)** Motifs from pGlu peptides in SAX fractions 1-8 (no motifs found for peptides in fractions 9-16). A higher score indicates a more significant/more specific motif. Fold increase compared to background of all sequences in the human proteome. Significance set at 1E-5 (not corrected for multiple hypotheses, corresponds to P value of 3E-4 with Bonferroni correction).

6.3 Conclusions

The original aim of this work was to develop phosphopeptide enrichment strategies for pHis analysis, but the unbiased SAX approach at non-acidic pH also lends itself to separation and characterisation of other acid-labile modifications. Additional MASCOT searches of this data revealed a vast number of previously unknown phosphorylation sites on Lys, Arg, Asp and Glu. The numbers of pLys, pArg, pAsp and pGlu peptides far exceed the number of pHis and pTyr peptides identified in these samples, meaning this unexplored area of the human phosphoproteome is perhaps more expansive than previously thought.

Interestingly, pLys phosphorylation was found to be increased in PHPT1 siRNA samples compared to NT siRNA treated controls, which is in direct contrast to the trend observed for all other sites of phosphorylation. Increased levels of pLys upon PHPT1 knockdown suggests that the pLys phosphatase activity of PHPT1 that was previously only observed *in vitro*[272], may also occur *in vivo*. The side chains of Lys and Arg are fully protonated at physiological pH (with pKa values of 10.4 and 12.3 respectively), resulting in these basic side chains carrying a net positive charge. The reversal of this charge upon addition of a negatively charged phosphate group will potentially enable switch-like functions, which may include regulation of catalytic outputs or protein binding activity.

Lys and Arg residues are also known to undergo various other PTMs[287], most notably methylation and acetylation, and phosphorylation of these residues may therefore provide a novel mechanism of competitive regulation. Given that Lys and Arg residues of histones are known to be phosphorylated[6, 41], and indeed three pLys peptides and one pArg peptide derived from histones were identified in this analysis, the balance between these modifications may have significant implications for understanding combinatorial signalling and epigenetics.

Glu and Asp residues possess a single negative charge at physiological pH, with side chain pKa values of 4.3 and 3.9 respectively. Addition of a phosphate group will result in a net negative charge >1 and a large hydrated ‘shell’ associated with the mixed anhydride structure. This marked increase in both the degree and directionality of the negative charge may have significant consequences for protein function, perhaps through induction of conformational changes to the protein structure.

Of particular interest in the pAsp data is the phosphorylation of PHPT1 on an Asp residue, with the same phosphopeptide identified in multiple samples. Further investigation of this pAsp site may reveal a possible mechanism by which the activity of PHPT1 is regulated. The phosphatase activity of PHPT1 has been reported to target both His and Lys, and indeed in this analysis the Lys phosphatase activity of PHPT1 seems particularly relevant. Therefore, it is conceivable that pAsp phosphorylation of PHPT1 may regulate levels of both pLys and pHis in human cells, hinting at the possibility of interplay between these non-canonical phosphorylation sites in cellular signalling events. To assess this hypothesis, in the first instance recombinant PHPT1 with or without mutation of this Asp residue to a non-phosphorylatable Asn, can be tested for pHis and pLys phosphatase activity *in vitro.*

Excitingly, this work highlights the prevalence of phosphorylation on residues other than Ser, Thr and Tyr, with insights already being revealed about their possible biological roles in human cells. Until now pLys, pArg, pAsp and pGlu were largely viewed as catalytic intermediates, but they also appear to occur as dynamic, stable PTMs in human cells. There is now a pressing need for new biochemical tools, such as phosphospecific antibodies, to further investigate these sites and build a more comprehensive picture of how they may function as part of signalling networks in mammalian systems.

Chapter 7. Discussion and Future Perspectives

The work in this thesis describes the characterisation of pHis and other acid-labile phosphorylated residues, using a workflow which incorporates a novel method termed UPAX (Unbiased Phosphopeptide enrichment by strong Anion Exchange). Implementation of the UPAX method combined with a comprehensive data analysis strategy for the identification and confident localisation of sites of phosphorylation, has revealed the vast extent of this previously unexplored area of the human phosphoproteome.

The first set of results presented in this thesis (*Chapter 3. Results I*) describes the assessment of various existing phosphopeptide enrichment techniques in terms of their suitability for recovery of phosphopeptides containing acid-labile pHis. A model protein, generated by chemical phosphorylation of myoglobin with potassium phosphoramidate, was digested with trypsin to yield five pHis-containing peptides. The stability of the pHis-containing peptides was assessed across a range of pH values and at elevated temperature, with $t_{1/2}$ defined under different conditions. At both pH 1 and 95 °C, $t_{1/2}$ of two representative pHis peptides was only 15 minutes, whilst at pH 4 and pH 6 pHis peptides remained phosphorylated for at least 2 hours. Given the acid-labile nature of pHis it was perhaps unsurprising that TiO_2 based enrichment which utilised acidic binding conditions was unsuccessful for pHis peptide enrichment. However, even with less acidic or near neutral buffers, recovery of pHis peptides from TiO_2 was still not achieved. The pH ~7 conditions utilised for hydroxyapatite enrichment were theoretically the most promising of all the techniques trialled, yet none of the model pHis peptides were recovered by this method. Furthermore, significant binding of non-phosphorylated peptides to the hydroxyapatite material was observed which is detrimental to overall enrichment of phosphopeptides. In summary, none of the peptide based enrichment methods tested for this work that had previously been used for phosphopeptide enrichment enabled recovery of pHis peptides.

In line with the most recent development in the field of pHis analysis, an immunoprecipitation method using 1- and 3-pHis monoclonal antibodies was utilised for protein level enrichment. Despite previously published work reporting selective enrichment of pHis proteins with the 1- and 3-pHis antibodies, neither that experiment or the one conducted for this work enabled detection of any pHis-containing peptides in the absence

of further peptide-based enrichment. Without observation of the relevant pHis peptide it is not possible to confirm that a protein is indeed phosphorylated. Furthermore, knowledge of specific pHis sites within a protein may reveal insight into possible functions which will not be gained by identification only at the protein level. It will be interesting in the future to perform the immunoprecipitation enrichment at the peptide level, a method which has proven successful for pTyr peptide enrichment using pTyr antibodies[166], or to combine the protein level enrichment with an additional peptide level enrichment step such as UPAX. It is likely that the reason pHis peptides are not detected following protein level immunoprecipitation is that digestion of these protein results in a complex mixture which contains a majority of non-phosphorylated peptides. Thus, an additional step to enrich and/or fractionate peptide samples may enable successful identification of pHis peptides by LC-MS/MS.

Following on from the unsuccessful attempts to enrich pHis-containing peptides using existing phosphopeptide enrichment strategies, the second results chapter (*Chapter 4. Results II*) describes the development of an alternative approach for separation and characterisation of pHis peptides based on SAX fractionation. A salt gradient elution using triethylammonium phosphate at pH 6.8 enabled separation of model pHis peptides from their non-phosphorylated equivalents. Phosphopeptides are more strongly retained by SAX as a result of their increased net negative charge, therefore fractions later in the gradient are enriched for phosphopeptides. For a simple mixture of peptides enrichment in later fractions reaches 100%, whilst for a complex cell lysate the fractions latest in gradient contain up to 50% phosphopeptides, representing a significant enrichment of phosphopeptides compared to an un-fractionated sample. Most importantly, the recovery of pHis-containing peptides by this SAX fractionation method was shown to be comparable to the level of recovery for other phosphorylated and non-phosphorylated peptides. Thus, the suitability of the UPAX method for characterisation of acid-labile pHis-containing peptides, even from complex mixtures, was established.

A key aspect of phosphopeptide characterisation is the confident localisation of the phosphate group to a specific site within the peptide sequence. Therefore, before embarking on further analysis of cell lysate samples this part of the data analysis pipeline was carefully considered. As detailed in *Chapter 5. Results III*, a potential diagnostic tool for confirming phosphorylation of His residues based on observation of characteristic neutral

loss ions in tandem MS spectra was evaluated. Essentially, the strategy proposed that triplet neutral loss of 80, 98 and 116 Da is characteristic of a pHis residue and thus could be used to confirm that a peptide is phosphorylated on His as opposed to another amino acid. However, the triplet neutral loss pattern was not consistently observed for peptides with confidently assigned pHis sites according to the ptmRS site localisation tool and/or manual annotation of spectra. Additionally, many phosphopeptides modified on other residues (Ser, Thr, Tyr) exhibited the triplet neutral loss, whilst an even greater number of phosphopeptide spectra did not contain any neutral loss ions following HCD fragmentation. Under the MS/MS conditions employed in these experiments, the triplet neutral loss strategy did not appear to be diagnostic for pHis peptides, and thus was not implemented for further data analysis.

The culmination of the method development was application of the UPAX strategy to HeLa cell lysate following treatment with PHPT1 siRNA or NT siRNA, and the subsequent identification of pHis and other acid-labile phosphorylation sites (pLys, pArg, pAsp and pGlu) from human cells (*Chapter 5. Results III* and *Chapter 6. Results IV)*. PHPT1 is a known histidine phosphatase and therefore it was hypothesised that a global increase in pHis levels would be observed following treatment of HeLa cells with PHPT1 siRNA. Contrary to this expectation, the number of phosphopeptides identified with each site of phosphorylation was increased in the NT siRNA treated samples compared to those treated with PHPT1 siRNA, with the exception of pLys-containing peptides. The increased level of pLys in PHPT1 siRNA treated cells aligns with the previously reported *in vitro* phosphatase activity of PHPT1 towards pLys[272], indicating that this Lys phosphatase activity may also be relevant *in vivo.* Additionally, the number of pLys, pArg, pAsp and pGlu sites identified across the two conditions far exceeds the number of pHis and pTyr sites identified in these samples, suggesting that this unexplored area of the phosphoproteome is far more extensive than previously understood.

One of the most exciting insights to be revealed by the results generated in this thesis is the phosphorylation of a conserved His residue within the catalytic domain of eukaryotic protein kinases. The relevant pHis site was observed in a number of peptides isolated from three different kinases, and was subsequently shown to map to a His residue conserved in 65% of eukaryotic protein kinases. The location of the HxN motif within the αC-β4 loop, which is implicated in positing of the αC-helix and thus catalytic activity, suggests

phosphorylation of the conserved His residue could serve as a general mechanism for kinase regulation, with results of MD simulations using the catalytic domain of PLK2 as a model system supporting this hypothesis. MD simulations suggest that addition of the 3-pHis isomer to His139 (the residue which maps to the HxN-His position) of PLK2 has a more pronounced effect than the 1-pHis isomer on structural changes within the activation segment, so a logical next step would be to determine which pHis isomer is present in PLK2. Fortunately, the availability of isomer specific monoclonal antibodies makes this feasible. PLK2 could be purified from cell extracts by immunoprecipitation with PLK2 antibodies, and subsequently immunoblotted with the 1- and 3-pHis antibodies, with the expectation that histidine phosphorylated PLK2 would only be detected by the antibody specific to either 1- or 3-pHis. Loss of immunoreactivity upon heat treatment will validate the presence of histidine phosphorylation, as opposed to phosphorylation of a heat stable residue or other non-specific binding of the antibody.

Further investigations into the consequences of phosphorylation of the H139 site in PLK2 may focus on mutation of the His reside to a residue that cannot be phosphorylated and does not carry a net charge, for example alanine. Comparison of the *in vitro* kinase activity of recombinantly expressed PLK2 and a H139A mutant, for example by radioactive kinase assay with [^{32}P]ATP and a known PLK2 substrate, may reveal how this pHis site is implicated in the catalytic activity of PLK2. Given the proposed destabilising effect of phosphorylation at the H139 site upon the αC-helix region, it might be expected that phosphorylation of the His residue will reduce (or perhaps completely abolish) kinase activity. Such studies using PLK2 as a model kinase will enable inferences to be made regarding a general mechanism of kinase regulation which is potentially relevant to the significant proportion of eukaryotic protein kinases containing this conserved His residue.

Given that PHPT1 is known to act as a histidine phosphatase, with activity demonstrated against ATP citrate lyase and the K^+ channels KCa3.1 and TRPV5[29-31], it was somewhat surprising that the expected reduction in pHis peptides upon treatment of HeLa cells with PHPT1 siRNA was not observed. It would therefore be interesting to understand which proteins, if any, of the ones identified in this work are direct targets for dephosphorylation by PHPT1. The difference in pHis peptide identifications between both conditions was quite pronounced, with a large portion of identifications unique to each one of the conditions, making identification of potential substrates of PHPT1 somewhat challenging based on this

data alone. Setting criteria to narrow down confidently identified pHis proteins from the PHPT1 siRNA treated samples, based on MASCOT and ptmRS scores for the pHis peptide, identification of other peptides from the protein, and observation across more than one bioreplicate, would be a good place to start in identifying potential substrates. To start to test for PHPT1 activity *in vitro,* phosphatase assays using recombinant PHPT1 against suspected protein substrates could be performed. Probing these samples with 1- and 3-pHis antibodies before and after treatment with PHPT1 would enable changes in histidine phosphorylation to be detected, thus identifying which substrates can be targeted for dephosphorylation by PHPT1.

PHPT1 has also been reported to act as a pLys phosphatase *in vitro*, and the increase in pLys peptides observed in this analysis upon treatment of HeLa cells with PHPT1 siRNA compared to NT siRNA, suggests that this activity may also be relevant *in vivo.* However, gaining insight into the pLys phosphatase activity of PHPT1 will be more challenging as there are no pLys antibodies, analogous to the monoclonal 1- and 3-pHis antibodies, currently reported. Understanding PHPT1 activity is further complicated by identification of a novel pAsp residue on PHPT1, which was observed across bioreplicates and has been confidently assigned by ptmRS score and manual annotation of tandem mass spectra. Unravelling how phosphorylation of both His and Lys may be regulated by pAsp presents an interesting direction for further study. The lack of appropriate biochemical tools to validate and further study non-canonical sites of phosphorylation may however be a significant impediment to the progression of this field of research. Although perhaps the unprecedented number of non-canonical phosphosites revealed by this work to be present in human cells will expedite development of new analytical strategies which will enable investigation of their roles in mammalian systems.

Alongside the potential development of new analytical approaches, increased certainty regarding the characterisation by LC-MS/MS of non-canonical sites of phosphorylation in human cells could be obtained by generation of high-resolution MS2 data, using the Orbitrap for mass analysis. The MS/MS method used for this work performed MS1 analysis in the Orbitrap and MS2 in the ion trap, with the rationale that the increased speed of MS2 ion trap analysis compared to the Orbitrap would result in more MS2 events and therefore more comprehensive coverage of peptides in the sample. This parallelization of MS1 and MS2 analysis has been reported to give the highest number of phosphopeptide

identifications, compared to methods which use the Orbitrap for both stages of MS analysis. However, in the same benchmarking study which was undertaken by someone else in the group while this data was being generated, optimal phosphosite localisation was achieved with Orbitrap MS2 analysis, with high mass accuracy increasing confidence in phosphopeptide identification and localisation compared to low resolution ion trap MS2 measurements[200]. Therefore, high resolution data generated by analysis of these samples using an Orbitrap for both MS1 and MS2 would increase the confidence associated with the assigned phosphosites, which is especially important when so many novel sites have been reported. Unfortunately, the samples generated for these experiments could not be re-analysed using an alternative LC-MS/MS method due to time limitations and problems of phosphopeptide instability upon even short-term storage. To enable further LC-MS/MS analysis repeat sample preparation would be required, including cell lysis, SAX fractionation and desalting. Based on the benchmarking study, high-resolution MS2 analysis would undoubtedly be a clear improvement to the analytical strategy presented in this thesis, should it be applied to future studies concerned with characterisation of the 'complete' phosphoproteome.

The data analysis pipeline used in this work to identify sites of non-canonical phosphorylation is arguably not the optimal approach. Searching each site of non-canonical phosphorylation separately does not take into account that some peptides may be phosphorylated on more than one of these non-canonical sites, and when considering only a single non-canonical pX at any one time there is potential for incorrect site assignment. It was observed that ~20% of all peptide spectrum matches (PSMs) containing non-canonical phosphorylation across all searches were derived from the same MS/MS scan, with a different pX site assigned as the site of phosphorylation at 1% FLR for the same peptide sequence. It is likely that the site of phosphorylation is incorrectly assigned in these peptides as the correct pX site was not offered as a potential site of modification. In the majority of cases where the site was differentially localised in different searches the possible sites of phosphorylation were within a few residues (<3) of each other within the peptide sequence, indicating that there are perhaps insufficient site-determining product ions to unambiguously localise the phosphosite.

A possible strategy to address the problem of the same MS/MS scans corresponding to different pX site assignments in different searches may be to extract these scans from the

data file and re-search them in MASCOT and ptmRS with all pX modifications considered, thus allowing the localisation algorithm to consider more than one non-canonical phosphorylation site for these peptides. The primary reason for not searching all of the data with all possible pX modifications is that by specifying many variable modifications for a MASCOT search, multiple versions of each peptide are created in the database, thus causing an exponential growth of the search space. Hence, not only is the run time of the search increased, but critically, the false discovery rate associated with the results alters dramatically. One potential way to include all pX modifications without drastically increasing the search space would be to reduce the protein database size. For example, an initial search of the data with just S, T and Y phosphorylation could be performed with proteins identified in this first search filtered at a stringent 1% protein FDR. The set of proteins that are confidently identified as being present in the sample would be much smaller than the entire UniProt database and thus would provide the database against which the data could be re-searched with all possible pX modifications included. The second search would produce a list of peptides for which all pX sites had been considered, for proteins that are most likely to be present in the sample, and thus a confident list of pX-containing peptides and proteins could be obtained.

There are also a number of software tools available for searching for many variable modifications, such as InsPecT[288], ModifiComb[289], and PeaksPTM[290], which perform unbiased PTM searches by a variety of different methods. These tools will search for multiple, undefined modifications, for example PeaksPTM searches the data with all modifications in the Uniprot database, not just the eight phosphosites of interest to this work. Although preliminary searching was attempted with PeaksPTM, it struggled to handle the mixed mode fragmentation workflow employed in these studies and total numbers of identified peptides (irrespective of phosphorylation) were low. Furthermore, site localisation confidence for the identified phosphosites is extremely important, meaning incorporation of results from unbiased searches into ptmRS would also require investigation. Attempting to implement alternative tools for PTM searching into the workflow for characterisation of non-canonical sites of phosphorylation would be a key step in the development of the data analysis pipeline for future applications. Indeed, there may be a requirement to develop new search strategies to assess the increased level of complexity within phosphoproteomics data afforded by the presence of so many possible sites of phosphorylation.

The combined UPAX, LC-MS/MS and data analysis strategy presented in this thesis has revealed widespread phosphorylation on a diverse range of amino acid residues other than Ser, Thr and Tyr in human systems. It appears that as well as the reported roles of these non-canonical phosphorylated residues as catalytic intermediates, they may also function as reversibly regulated PTMs in cellular signalling pathways. Further understanding of the possible roles of these non-canonical phosphorylation sites in cellular signalling will require the kinases and phosphatases responsible for their modulation to be ascertained, which is a considerable biochemical challenge. The regulation of these phosphosites as part of complex signalling networks may also be understood by revealing common protein domains on which these sites are observed. Molecular biology and chemical genetics strategies will further aid in elucidation of the mechanisms by which these phosphosites are implicated in cellular functions, whilst evaluation of non-canonical phosphorylation in the context of the cell cycle, ageing, disease models and drug treatments will all contribute to the understanding of the roles of these phosphorylated residues in global cell biology. Furthermore, the strategy presented in this thesis will also be applicable to analysis of non-canonical phosphorylation in other systems, perhaps enabling large-scale characterisation of these sites in bacteria, plants and other eukaryotes, which are already known to use pHis, pAsp and pArg for signal transduction[11, 12, 16, 234]. The relative ease with which the UPAX strategy alongside LC-MS/MS analysis can be utilised for characterisation of these non-canonical phosphorylation sites in a complex mixture will undoubtedly be instrumental in enabling their extent and functional relevance in a wide variety of biological systems to be revealed, thus facilitating significant changes in the way we currently understand phosphorylation-mediated signalling.

Chapter 8. References

1. Adams, J. A., Kinetic and catalytic mechanisms of protein kinases, *Chem. Rev.*, **2001**, 101 (8), 2271-2290.
2. Levene, P. A. and Alsberg, C. L., The cleavage products of Vitellin, *J. Biol. Chem.*, **1906**, 2 (1), 127-133.
3. Burnett, G. and Kennedy, E. P., The enzymatic phosphorylation of proteins, *J. Biol. Chem.*, **1954**, 211 (2), 969-980.
4. Cohen, P., The role of protein phosphorylation in human health and disease., *Eur. J. Biochem.*, **2001**, 268 (19), 5001-5010.
5. Sickmann, A. and Meyer, H. E., Phosphoamino acid analysis, *Proteomics*, **2001**, 1 (2), 200-206.
6. Besant, P. G., Attwood, P. V.*, et al.*, Focus on phosphoarginine and phospholysine, *Curr. Protein Pept. Sci.*, **2009**, 10 (6), 536-550.
7. Attwood, P. V., Besant, P. G.*, et al.*, Focus on phosphoaspartate and phosphoglutamate, *Amino Acids*, **2011**, 40 (4), 1035-1051.
8. Stadtman, T. C., Emerging awareness of the critical roles of S-phosphocysteine and selenophosphate in biological systems, *BioFactors (Oxford, England)*, **1994**, 4 (3-4), 181-185.
9. Pesis, K. H., Wei, Y. F.*, et al.*, Phosphohistidine is found in basic nuclear proteins of physarum-polycephalum, *Febs Lett.*, **1988**, 239 (1), 151-154.
10. Boyer, P. D., Deluca, M.*, et al.*, Identification of phosphohistidine in digests from a probable intermediate of oxidative phosphorylation, *J. Biol. Chem.*, **1962**, 237, Pc3306-pc3308.
11. Stock, A. M., Robinson, V. L.*, et al.*, Two-component signal transduction, *Annu. Rev. Biochem.*, **2000**, 69, 183-215.
12. Saito, H., Histidine phosphorylation and two-component signaling in eukaryotic cells *Chem. Rev.*, **2001**, 101, 2497-2509.
13. Besant, P. G., Tan, E.*, et al.*, Mammalian protein histidine kinases, *Int. J. Biochem. Cell Biol.*, **2003**, 35 (3), 297-309.
14. Mizuno, T., His-Asp phosphotransfer signal transduction, *J. Biochem.*, **1998**, 123 (4), 555-563.
15. Nakamichi, N., Yamada, H.*, et al.*, His-to-Asp Phosphorelay Circuitry for Regulation of Sexual Development in Schizosaccharomyces pombe, *Biosci. Biotechnol. Biochem.*, **2002**, 66 (12), 2663-2672.
16. Mizuno, T., Two-component phosphorelay signal transduction systems in plants: from hormone responses to circadian rhythms, *Biosci. Biotechnol. Biochem.*, **2005**, 69 (12), 2263-2276.
17. Attwood, P. V., Histidine kinases from bacteria to humans, *Biochem. Soc. T.*, **2013**, 41, 1023-1028.
18. Rose, Z. B., Evidence for a phosphohistidine protein intermediate in the phosphoglycerate mutase reaction, *Arch. Biochem. Biophys.*, **1970**, 140 (2), 508-513.
19. Gottlin, E. B., Rudolph, A. E.*, et al.*, Catalytic mechanism of the phospholipase D superfamily proceeds via a covalent phosphohistidine intermediate, *Proc. Natl. Acad. Sci. USA*, **1998**, 95 (16), 9202-9207.
20. Ghosh, A., Shieh, J.-J.*, et al.*, Histidine 167 Is the Phosphate Acceptor in Glucose-6-phosphatase-β Forming a Phosphohistidine Enzyme Intermediate during Catalysis, *J. Biol. Chem.*, **2004**, 279 (13), 12479-12483.

21. Thomas, D. H., Getz, T. M.*, et al.*, A novel histidine tyrosine phosphatase, TULA-2, associates with Syk and negatively regulates GPVI signaling in platelets, *Blood*, **2010**, 116 (14), 2570-2578.
22. Back, S. H., Adapala, N. S.*, et al.*, TULA-2, a novel histidine phosphatase, regulates bone remodeling by modulating osteoclast function, *Cell Mol. Life Sci.*, **2013**, 70 (7), 1269-1284.
23. Matthews, H. R., Protein kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins: A possible regulator of the mitogen-activated protein kinase cascade, *Pharmacology & Therapeutics*, **1995**, 67 (3), 323-350.
24. Hegde, A. N., Swamy, C. V.*, et al.*, Negative correlation with liver cell division of a 38 kilodalton protein whose phosphorylation is enhanced by ras and G-proteins, *FEBS Lett.*, **1993**, 333 (1-2), 103-107.
25. Steeg, P. S., Bevilacqua, G.*, et al.*, Evidence for a novel gene associated with low tumor metastatic potential, *J. Natl. Cancer Inst.*, **1988**, 80 (3), 200-204.
26. Hartsough, M. T. and Steeg, P. S., Nm23/nucleoside diphosphate kinase in human cancers, *J. Bioenerg. Biomembr.*, **2000**, 32 (3), 301-308.
27. Thakur, R. K., Yadav, V. K.*, et al.*, Mechanisms of non-metastatic 2 (NME2)-mediated control of metastasis across tumor types, *N-S Arch. Pharmacol.*, **2011**, 384 (4-5), 397-406.
28. Cuello, F., Schulze, R. A.*, et al.*, Activation of heterotrimeric G proteins by a high energy phosphate transfer via nucleoside diphosphate kinase (NDPK) B and Gbeta subunits. Complex formation of NDPK B with Gbeta gamma dimers and phosphorylation of His-266 IN Gbeta, *J. Biol. Chem.*, **2003**, 278 (9), 7220-7226.
29. Wagner, P. D. and Vu, N.-D., Phosphorylation of ATP-Citrate Lyase by Nucleoside Diphosphate Kinase, *J. Biol. Chem.*, **1995**, 270 (37), 21758-21764.
30. Srivastava, S., Li, Z.*, et al.*, Histidine phosphorylation of the potassium channel KCa3.1 by nucleoside diphosphate kinase B is required for activation of KCa3.1 and CD4 T cells, *Mol. Cell*, **2006**, 24 (5), 665-675.
31. Cai, X., Srivastava, S.*, et al.*, Regulation of the epithelial Ca2+ channel TRPV5 by reversible histidine phosphorylation mediated by NDPK-B and PHPT1, *Mol. Biol. Cell*, **2014**, 25 (8), 1244-1250.
32. Ek, P., Pettersson, G.*, et al.*, Identification and characterization of a mammalian 14-kDa phosphohistidine phosphatase, *Eur. J. Biochem.*, **2002**, 269 (20), 5016-5023.
33. Klumpp, S., Hermesmeier, J.*, et al.*, Protein histidine phosphatase: a novel enzyme with potency for neuronal signaling, *J. Cereb. Blood Flow Metab.*, **2002**, 22 (12), 1420-1424.
34. Srivastava, S., Zhdanova, O.*, et al.*, Protein histidine phosphatase 1 negatively regulates CD4 T cells by inhibiting the K+ channel KCa3.1, *Proc. Natl. Acad. Sci. USA*, **2008**, 105 (38), 14442-14446.
35. Panda, S., Srivastava, S.*, et al.*, Identification of PGAM5 as a Mammalian Protein Histidine Phosphatase that Plays a Central Role to Negatively Regulate CD4+ T Cells, *Mol. Cell*, **2016**, 63 (3), 457-469.
36. Crovello, C. S., Furie, B. C.*, et al.*, Histidine phosphorylation of P-selectin upon stimulation of human platelets: A novel pathway for activation-dependent signal transduction, *Cell*, **1995**, 82 (2), 279-286.
37. Muimo, R., Hornickova, Z.*, et al.*, Histidine phosphorylation of annexin I in airway epithelia, *J. Biol. Chem.*, **2000**, 275 (47), 36632-36636.
38. Fujitaki, J. M., Fung, G.*, et al.*, Characterization of chemical and enzymatic acid-labile phosphorylation of histone H4 using phosphorus-31 nuclear magnetic resonance, *Biochemistry*, **1981**, 20 (12), 3658-3664.

39. Smtih, D. L., Bruegger, B. B.*, et al.*, New histone kinases in nuclei of rat tissues, *Nature*, **1973**, 246 (5428), 103-104.
40. Smith, D. L., Chen, C.-C.*, et al.*, Characterization of protein kinases forming acid-labile histone phosphates in Walker-256 carcinosarcoma cell nuclei, *Biochemistry*, **1974**, 13 (18), 3780-3785.
41. Chen, C.-C., Smith, D. L.*, et al.*, Occurrence and distribution of acid-labile histone phosphates in regenerating rat liver, *Biochemistry*, **1974**, 13 (18), 3785-3789.
42. Tan, E., Besant, P. G.*, et al.*, Histone H4 histidine kinase displays the expression pattern of a liver oncodevelopmental marker, *Carcinogenesis*, **2004**, 25 (11), 2083-2088.
43. Besant, P. G. and Attwood, P. V., Histone H4 histidine phosphorylation: kinases, phosphatases, liver regeneration and cancer, *Biochem. Soc. T.*, **2012**, 40 (1), 290-293.
44. Wilkins, M. R., Pasquali, C.*, et al.*, From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis, *Bio/technology (Nature Publishing Company)*, **1996**, 14 (1), 61-65.
45. de Hoog, C. L. and Mann, M., Proteomics, *Annu. Rev. Genomics Hum. Genet.*, **2004**, 5, 267-293.
46. Gygi, S. P., Corthals, G. L.*, et al.*, Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, *Proc. Natl. Acad. Sci. USA*, **2000**, 97 (17), 9390-9395.
47. Lilley, K. S., Razzaq, A.*, et al.*, Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation, *Curr. Opin. Chem. Biol.*, **2002**, 6 (1), 46-50.
48. Meng, C. K., Mann, M.*, et al.*, Of protons or proteins, *Z. Phys. D Atoms, Molecules and Clusters*, **1988**, 10 (2), 361-368.
49. Fenn, J. B., Mann, M.*, et al.*, Electrospray ionization for mass spectrometry of large biomolecules, *Science*, **1989**, 246 (4926), 64-71.
50. Wilm, M. S. and Mann, M., Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?, *Int. J. Mass Spectrom. Ion Process.*, **1994**, 136 (2), 167-180.
51. Wilm, M. and Mann, M., Analytical Properties of the Nanoelectrospray Ion Source, *Anal. Chem.*, **1996**, 68 (1), 1-8.
52. Iribarne, J. V. and Thomson, B. A., On the evaporation of small ions from charged droplets, *J. Chem. Phys.*, **1976**, 64 (6), 2287-2294.
53. Thomson, B. A. and Iribarne, J. V., Field induced ion evaporation from liquid surfaces at atmospheric pressure, *J. Chem. Phys.*, **1979**, 71 (11), 4451-4463.
54. Dole, M., Mack, L. L.*, et al.*, Molecular Beams of Macroions, *J. Chem. Phys.*, **1968**, 49 (5), 2240-2249.
55. Wilm, M., Principles of electrospray ionization, *Molecular & Cellular Proteomics*, **2011**.
56. Paul, W. and Steinwedel, H., *U.S Patent 2,939,952*, **1960**.
57. Stafford, G. C., Kelley, P. E.*, et al.*, Recent improvements in and analytical applications of advanced ion trap technology, *Int. J. Mass Spectrom. Ion Process.*, **1984**, 60 (1), 85-98.
58. Todd, J. F. J., Ion trap mass spectrometer - past, present and future, *Mass Spectrom. Rev.*, **1991**, 10 (1), 3-52.
59. Douglas, D. J., Frank, A. J.*, et al.*, Linear ion traps in mass spectrometry, *Mass Spectrom. Rev.*, **2005**, 24 (1), 1-29.
60. Schwartz, J. C., Senko, M. W.*, et al.*, A two-dimensional quadrupole ion trap mass spectrometer, *J. Am. Soc. Mass Spectrom.*, **2002**, 13 (6), 659-669.

61. Schwartz, J. C., Zhou, X. G.*, et al.*, *U.S*, **1996**.
62. Makarov, A. A., *U.S.*, **1999**.
63. Makarov, A., Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis, *Anal. Chem.*, **2000**, 72 (6), 1156-1162.
64. Dongré, A. R., Jones, J. L.*, et al.*, Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model, *J. Am. Chem. Soc.*, **1996**, 118 (35), 8365-8374.
65. Summerfield, S. G. and Gaskell, S. J., Fragmentation efficiencies of peptide ions following low energy collisional activation, *Int. J. Mass Spectrom. Ion Process.*, **1997**, 165, 509-521.
66. Wysocki, V. H., Tsaprailis, G.*, et al.*, Mobile and localized protons: a framework for understanding peptide dissociation, *J. Mass Spectrom.*, **2000**, 35 (12), 1399-1406.
67. Roepstorff, P. and Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides, *Biol. Mass Spectrom.*, **1984**, 11 (11), 601-601.
68. Biemann, K., Contributions of mass spectrometry to peptide and protein structure, *Biol. Mass Spectrom.*, **1988**, 16 (1-12), 99-111.
69. Olsen, J. V., Macek, B.*, et al.*, Higher-energy C-trap dissociation for peptide modification analysis, *Nat. Methods*, **2007**, 4 (9), 709-712.
70. Lau, K. W., Hart, S. R.*, et al.*, Observations on the detection of b- and y-type ions in the collisionally activated decomposition spectra of protonated peptides, *Rapid Commun. Mass Spectrom.*, **2009**, 23 (10), 1508-1514.
71. Syka, J. E., Coon, J. J.*, et al.*, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, *Proc. Natl. Acad. Sci. USA*, **2004**, 101 (26), 9528-9533.
72. Mikesh, L. M., Ueberheide, B.*, et al.*, The Utility of ETD Mass Spectrometry in Proteomic Analysis, *Biochim. Biophys. Acta.*, **2006**, 1764 (12), 1811-1822.
73. Good, D. M., Wirtala, M.*, et al.*, Performance characteristics of electron transfer dissociation mass spectrometry, *Mol. Cell. Proteomics*, **2007**, 6 (11), 1942-1951.
74. Molina, H., Matthiesen, R.*, et al.*, Comprehensive Comparison of Collision Induced Dissociation and Electron Transfer Dissociation, *Anal. Chem.*, **2008**, 80 (13), 4825-4835.
75. Frese, C. K., Altelaar, A. F.*, et al.*, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos, *J. Proteome Res.*, **2011**, 10 (5), 2377-2388.
76. Swaney, D. L., McAlister, G. C.*, et al.*, Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors, *Anal. Chem.*, **2007**, 79 (2), 477-485.
77. Frese, C. K., Altelaar, A. F. M.*, et al.*, 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.
78. Frese, C. K., Zhou, H.*, et al.*, Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD), *J. Proteome Res.*, **2013**, 12 (3), 1520-1525.
79. Liu, F., van Breukelen, B.*, et al.*, Facilitating protein disulfide mapping by a combination of pepsin digestion, electron transfer higher energy dissociation (EThcD), and a dedicated search algorithm SlinkS, *Mol. Cell. Proteomics*, **2014**, 13 (10), 2776-2786.
80. Brunner, A. M., Lossl, P.*, et al.*, Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization, *Anal. Chem.*, **2015**, 87 (8), 4152-4158.

81. Bilan, V., Leutert, M.*, et al.*, Combining Higher-Energy Collision Dissociation and Electron-Transfer/Higher-Energy Collision Dissociation Fragmentation in a Product-Dependent Manner Confidently Assigns Proteomewide ADP-Ribose Acceptor Sites, **2017**, 89 (3), 1523-1530.
82. Senko, M. W., Remes, P. M.*, et al.*, Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates, *Anal. Chem.*, **2013**, 85 (24), 11710-11714.
83. Eliuk, S. and Makarov, A., Evolution of Orbitrap Mass Spectrometry Instrumentation, *Annu. Rev. Anal. Chem.*, **2015**, 8, 61-80.
84. Washburn, M. P., Wolters, D.*, et al.*, Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat Biotechnol*, **2001**, 19 (3), 242-247.
85. Stekhoven, D. J., Omasits, U.*, et al.*, Proteome-wide identification of predominant subcellular protein localizations in a bacterial model organism, *J. Proteomics*, **2014**, 99, 123-137.
86. Li, Z., Czarnecki, O.*, et al.*, Strigolactone-Regulated Proteins Revealed by iTRAQ-Based Quantitative Proteomics in Arabidopsis, *J. Proteome Res.*, **2014**, 13 (3), 1359-1372.
87. Kim, M. S., Pinto, S. M.*, et al.*, A draft map of the human proteome, *Nature*, **2014**, 509 (7502), 575-581.
88. Feng, Y., De Franceschi, G.*, et al.*, Global analysis of protein structural changes in complex proteomes, *Nat. Biotechnol.*, **2014**, 32 (10), 1036-1044.
89. Rigaut, G., Shevchenko, A.*, et al.*, A generic protein purification method for protein complex characterization and proteome exploration, *Nat. Biotechnol.*, **1999**, 17 (10), 1030-1032.
90. Larance, M., Kirkwood, K. J.*, et al.*, Global Membrane Protein Interactome Analysis using In vivo Crosslinking and Mass Spectrometry-based Protein Correlation Profiling, *Mol. Cell. Proteomics*, **2016**, 15 (7), 2476-2490.
91. Dunkley, T. P., Watson, R.*, et al.*, Localization of organelle proteins by isotope tagging (LOPIT), *Mol. Cell. Proteomics*, **2004**, 3 (11), 1128-1134.
92. Huttlin, E. L., Jedrychowski, M. P.*, et al.*, A tissue-specific atlas of mouse protein phosphorylation and expression, *Cell*, **2010**, 143 (7), 1174-1189.
93. Chicooree, N., Griffiths, J. R.*, et al.*, A novel approach to the analysis of SUMOylation with the independent use of trypsin and elastase digestion followed by database searching utilising consecutive residue addition to lysine, *Rapid Commun. Mass Spectrom.*, **2013**, 27 (1), 127-134.
94. Zhou, H., Di Palma, S.*, et al.*, Toward a Comprehensive Characterization of a Human Cancer Cell Phosphoproteome, *J. Proteome Res.*, **2013**, 12 (1), 260-271.
95. Martello, R., Leutert, M.*, et al.*, Proteome-wide identification of the endogenous ADP-ribosylome of mammalian cells and tissue, *Nat. Commun.*, **2016**, 7, 12917.
96. Siuti, N., Roth, M. J.*, et al.*, Gene-specific characterization of human histone H2B by electron capture dissociation, *J. Proteome Res.*, **2006**, 5 (2), 233-239.
97. Barrera, N. P., Isaacson, S. C.*, et al.*, Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions, *Nat. Methods*, **2009**, 6 (8), 585-587.
98. Zhang, J., Guy, M. J.*, et al.*, Top-down quantitative proteomics identified phosphorylation of cardiac troponin I as a candidate biomarker for chronic heart failure, *J. Proteome Res.*, **2011**, 10 (9), 4054-4065.
99. Savaryn, J. P., Toby, T. K.*, et al.*, Comparative top down proteomics of peripheral blood mononuclear cells from kidney transplant recipients with normal kidney biopsies or acute rejection, *Proteomics*, **2016**, 16 (14), 2048-2058.

100. Liu, H., Sadygov, R. G.*, et al.*, A model for random sampling and estimation of relative protein abundance in shotgun proteomics, *Anal. Chem.*, **2004**, 76 (14), 4193-4201.
101. Chelius, D. and Bondarenko, P. V., Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry, *J. Proteome Res.*, **2002**, 1 (4), 317-323.
102. Silva, J. C., Gorenstein, M. V.*, et al.*, Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition, *Mol. Cell. Proteomics*, **2006**, 5 (1), 144-156.
103. Ong, S. E., Blagoev, B.*, et al.*, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, *Mol. Cell. Proteomics*, **2002**, 1 (5), 376-386.
104. Thompson, A., Schäfer, J.*, et al.*, Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS, *Anal. Chem.*, **2003**, 75 (8), 1895-1904.
105. Ross, P. L., Huang, Y. N.*, et al.*, Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents, *Mol. Cell. Proteomics*, **2004**, 3 (12), 1154-1169.
106. Silva, J. C., Denny, R.*, et al.*, Quantitative proteomic analysis by accurate mass retention time pairs, *Anal Chem*, **2005**, 77 (7), 2187-2200.
107. Gillet, L. C., Navarro, P.*, et al.*, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, *Molecular & cellular proteomics : MCP*, **2012**, 11 (6), O111.016717.
108. Jensen, O. N., Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry, *Curr. Opin. Chem. Biol.*, **2004**, 8 (1), 33-41.
109. Liao, P. C., Leykam, J.*, et al.*, An Approach to Locate Phosphorylation Sites in a Phosphoprotein: Mass Mapping by Combining Specific Enzymatic Degradation with Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry, *Anal. Biochem.*, **1994**, 219 (1), 9-20.
110. Kussmann, M., Hauser, K.*, et al.*, Comparison of in vivo and in vitro phosphorylation of the exocytosis-sensitive protein PP63/parafusin by differential MALDI mass spectrometric peptide mapping, *Biochemistry*, **1999**, 38 (24), 7780-7790.
111. Larsen, M. R., Sorensen, G. L.*, et al.*, Phospho-proteomics: evaluation of the use of enzymatic de-phosphorylation and differential mass spectrometric peptide mass mapping for site specific phosphorylation assignment in proteins separated by gel electrophoresis, *Proteomics*, **2001**, 1 (2), 223-238.
112. Covey, T., Shushan, B.*, et al.*, LC/MS and LC/MS/MS Screening for the Sites of Post-Translational Modification in Proteins in *Methods in Protein Sequence Analysis*, eds. Jörnvall, H., Höög, J.-O.*, et al.*, Birkhäuser Basel, **1991**, pp. 249-256.
113. Carr, S. A., Huddleston, M. J.*, et al.*, Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry, *Anal. Biochem.*, **1996**, 239 (2), 180-192.
114. Wilm, M., Neubauer, G.*, et al.*, Parent Ion Scans of Unseparated Peptide Mixtures, *Anal. Chem.*, **1996**, 68 (3), 527-533.
115. MacCoss, M. J., McDonald, W. H.*, et al.*, Shotgun identification of protein modifications from protein complexes and lens tissue, *Proc. Natl. Acad. Sci. USA*, **2002**, 99 (12), 7900-7905.
116. Ficarro, S. B., McCleland, M. L.*, et al.*, Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae, *Nat. Biotechnol.*, **2002**, 20 (3), 301-305.

117. Olsen, J. V., Blagoev, B.*, et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks, *Cell*, **2006**, 127 (3), 635-648.
118. Lanucara, F., Lam, C.*, et al.*, Dynamic phosphorylation of RelA on Ser42 and Ser45 in response to TNFα stimulation regulates DNA binding and transcription, *Open Biol.*, **2016**, 6 (7), 160055.
119. Harsha, H. C. and Pandey, A., Phosphoproteomics in cancer, *Mol. Oncol.*, **2010**, 4 (6), 482-495.
120. Neville, D. C., Rozanas, C. R.*, et al.*, Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry, *Protein Sci.*, **1997**, 6 (11), 2436-2445.
121. Posewitz, M. C. and Tempst, P., Immobilized gallium(III) affinity chromatography of phosphopeptides, *Anal. Chem.*, **1999**, 71 (14), 2883-2892.
122. Negroni, L., Claverol, S.*, et al.*, Comparison of IMAC and MOAC for phosphopeptide enrichment by column chromatography, *J. Chromatogr. B.*, **2012**, 891, 109-112.
123. Stewart, I. I., Thomson, T.*, et al.*, 18O Labeling: a tool for proteomics, *Rapid Commun. Mass Spectrom.*, **2001**, 15 (24), 2456-2465.
124. Jensen, S. S. and Larsen, M. R., Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques, *Rapid Commun. Mass Spectrom.*, **2007**, 21 (22), 3635-3645.
125. McNulty, D. E. and Annan, R. S., Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection, *Mol. Cell. Proteomics*, **2008**, 7 (5), 971-980.
126. Trinidad, J. C., Specht, C. G.*, et al.*, Comprehensive Identification of Phosphorylation Sites in Postsynaptic Density Preparations, *Mol. Cell. Proteomics*, **2006**, 5 (5), 914-922.
127. Villen, J. and Gygi, S. P., The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry, *Nat. Protocols*, **2008**, 3 (10), 1630-1638.
128. Zhou, H., Tian, R.*, et al.*, Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis, *Electrophoresis*, **2007**, 28 (13), 2201-2215.
129. Li, Y., Liu, Y.*, et al.*, Fe3O4@Al2O3 magnetic core–shell microspheres for rapid and highly specific capture of phosphopeptides with mass spectrometry analysis, *J. Chromatogr. A*, **2007**, 1172 (1), 57-71.
130. Ficarro, S. B., Parikh, J. R.*, et al.*, Niobium(V) Oxide (Nb2O5): Application to Phosphoproteomics, *Anal. Chem.*, **2008**, 80 (12), 4606-4613.
131. Lu, J., Qi, D.*, et al.*, Hydrothermal synthesis of alpha-Fe(2)O(3)@SnO(2) core-shell nanotubes for highly selective enrichment of phosphopeptides for mass spectrometry analysis, *Nanoscale*, **2010**, 2 (10), 1892-1900.
132. Sun, S., Ma, H.*, et al.*, Efficient enrichment and identification of phosphopeptides by cerium oxide using on-plate matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis, *Rapid Commun. Mass Spectrom.*, **2011**, 25 (13), 1862-1868.
133. Sano, A. and Nakamura, H., Titania as a Chemo-affinity Support for the Column-switching HPLC Analysis of Phosphopeptides: Application to the Characterization of Phosphorylation Sites in Proteins by Combination with Protease Digestion and Electrospray Ionization Mass Spectrometry, *Analytical Sciences*, **2004**, 20 (5), 861-864.
134. Kuroda, I., Shintani, Y.*, et al.*, Phosphopeptide-selective Column-switching RP-HPLC with a Titania Precolumn, *Analytical Sciences*, **2004**, 20 (9), 1313-1319.

135. Pinkse, M. W. H., Uitto, P. M.*, et al.*, Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns, *Anal. Chem.*, **2004**, 76 (14), 3935-3943.
136. Larsen, M. R., Thingholm, T. E.*, et al.*, Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns, *Mol. Cell. Proteomics*, **2005**, 4 (7), 873-886.
137. Thingholm, T. E., Jorgensen, T. J. D.*, et al.*, Highly selective enrichment of phosphorylated peptides using titanium dioxide, *Nat. Protoc.*, **2006**, 1 (4), 1929-1935.
138. Wu, J., Shakey, Q.*, et al.*, Global profiling of phosphopeptides by titania affinity enrichment, *J. Proteome Res.*, **2007**, 6 (12), 4684-4689.
139. Sugiyama, N., Masuda, T.*, et al.*, Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications, *Mol. Cell. Proteomics*, **2007**, 6 (6), 1103-1109.
140. Fukuda, I., Hirabayashi-Ishioka, Y.*, et al.*, Optimization of enrichment conditions on TiO2 chromatography using glycerol as an additive reagent for effective phosphoproteomic analysis, *J. Proteome Res.*, **2013**, 12 (12), 5587-5597.
141. Vilasi, A., Fiume, I.*, et al.*, Enrichment specificity of micro and nano-sized titanium and zirconium dioxides particles in phosphopeptide mapping, *J. Mass Spectrom.*, **2013**, 48 (11), 1188-1198.
142. Zhang, L., Liang, Z.*, et al.*, Mesoporous TiO2 aerogel for selective enrichment of phosphopeptides in rat liver mitochondria, *Anal. Chim. Acta*, **2012**, 729, 26-35.
143. Wang, S. T., Wang, M. Y.*, et al.*, Facile preparation of SiO2/TiO2 composite monolithic capillary column and its application in enrichment of phosphopeptides, *Anal. Chem.*, **2012**, 84 (18), 7763-7770.
144. Zeng, Y. Y., Chen, H. J.*, et al.*, Efficient enrichment of phosphopeptides by magnetic TiO(2)-coated carbon-encapsulated iron nanoparticles, *Proteomics*, **2012**, 12 (3), 380-390.
145. Bi, H., Qiao, L.*, et al.*, TiO(2) printed aluminum foil: single-use film for a laser desorption/ionization target plate, *Anal. Chem.*, **2009**, 81 (3), 1177-1183.
146. Chen, C.-J., Lai, C.-C.*, et al.*, A novel titanium dioxide-polydimethylsiloxane plate for phosphopeptide enrichment and mass spectrometry analysis, *Anal. Chim. Acta*, **2014**, 812 (0), 105-113.
147. Engholm-Keller, K., Hansen, T. A.*, et al.*, Multidimensional strategy for sensitive phosphoproteomics incorporating protein prefractionation combined with SIMAC, HILIC, and TiO(2) chromatography applied to proximal EGF signaling, *J. Proteome Res.*, **2011**, 10 (12), 5383-5397.
148. Li, Q.-R., Ning, Z.-B.*, et al.*, Complementary workflow for global phosphoproteome analysis, *Electrophoresis*, **2012**, 33 (22), 3291-3298.
149. Herring, L. E., Grant, K. G.*, et al.*, Development of a tandem affinity phosphoproteomic method with motif selectivity and its application in analysis of signal transduction networks, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, **2015**, 988, 166-174.
150. Gorbunoff, M. J., [26] Protein chromatography on hydroxyapatite columns in *Methods in enzymology*, ed. Murray, P. D., Academic Press, **1990**, pp. 329-339.
151. Bernardi, G., Chromatography of nucleic acids on hydroxyapatite I. Chromatography of native DNA, *BBA-Nucleic Acids and Protein Synthesis*, **1969**, 174 (2), 423-434.
152. Gagnon, P., Monoclonal antibody purification with hydroxyapatite, *New Biotechnol.*, **2009**, 25 (5), 287-293.

153. Hussey, C. R., Liddle, P. F.*, et al.*, The isolation and characterization of differentially phosphorylated fractions of phosphofructokinase from rabbit skeletal muscle, *Eur. J. Biochem.*, **1977**, 80 (2), 497-506.
154. Tashima, Y., Terui, M.*, et al.*, Phosphorylated and dephosphorylated types of non-activated glucocorticoid receptor, *J. Biochem.*, **1990**, 108 (2), 271-277.
155. Mamone, G., Picariello, G.*, et al.*, Hydroxyapatite affinity chromatography for the highly selective enrichment of mono- and multi-phosphorylated peptides in phosphoproteome analysis, *Proteomics*, **2010**, 10 (3), 380-393.
156. Pinto, G., Caira, S.*, et al.*, Hydroxyapatite as a concentrating probe for phosphoproteomic analyses, *J. Chromatogr. B.*, **2010**, 878 (28), 2669-2678.
157. Krenkova, J. and Foret, F., Nanoparticle-modified monolithic pipette tips for phosphopeptide enrichment, *Anal Bioanal Chem*, **2013**, 405 (7), 2175-2183.
158. Yu, Q., Li, X.-S.*, et al.*, Preparation of magnetic hydroxyapatite clusters and their application in the enrichment of phosphopeptides, *J. Sep. Sci.*, **2014**, 37 (5), 580-586.
159. Fonslow, B. R., Niessen, S. M.*, et al.*, Single-step inline hydroxyapatite enrichment facilitates identification and quantitation of phosphopeptides from mass-limited proteomes with MudPIT, *J. Proteome Res.*, **2012**, 11 (5), 2697-2709.
160. Zhang, X., Ye, J.*, et al.*, Highly efficient phosphopeptide enrichment by calcium phosphate precipitation combined with subsequent IMAC enrichment, *Mol. Cell. Proteomics*, **2007**, 6 (11), 2032-2042.
161. Xia, Q. W., Cheng, D. M.*, et al.*, Phosphoproteomic analysis of human brain by calcium phosphate precipitation and mass spectrometry, *J. Proteome Res.*, **2008**, 7 (7), 2845-2851.
162. Mirza, M. R., Rainer, M.*, et al.*, A novel strategy for phosphopeptide enrichment using lanthanide phosphate co-precipitation, *Anal Bioanal Chem*, **2012**, 404 (3), 853-862.
163. Pandey, A., Podtelejnikov, A. V.*, et al.*, Analysis of receptor signaling pathways by mass spectrometry: Identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors, *Proc. Natl. Acad. Sci. USA*, **2000**, 97 (1), 179-184.
164. Salomon, A. R., Ficarro, S. B.*, et al.*, Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry, *Proc. Natl. Acad. Sci. USA*, **2003**, 100 (2), 443-448.
165. Brill, L. M., Salomon, A. R.*, et al.*, Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry, *Anal. Chem.*, **2004**, 76 (10), 2763-2772.
166. Rush, J., Moritz, A.*, et al.*, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells, *Nat. Biotechnol.*, **2005**, 23 (1), 94-101.
167. Rikova, K., Guo, A.*, et al.*, Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer, **2007**, 131 (6), 1190-1203.
168. Gronborg, M., Kristiansen, T. Z.*, et al.*, A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate, *Mol. Cell. Proteomics*, **2002**, 1 (7), 517-527.
169. Shen, Y., Jacobs, J. M.*, et al.*, Ultra-High-Efficiency Strong Cation Exchange LC/RPLC/MS/MS for High Dynamic Range Characterization of the Human Plasma Proteome, *Anal. Chem.*, **2004**, 76 (4), 1134-1144.
170. Chiu, C.-W., Chang, C.-L.*, et al.*, Evaluation of peptide fractionation strategies used in proteome analysis, *J. Sep. Sci.*, **2012**, 35 (23), 3293-3301.

171. Gilar, M., Olivova, P.*, et al.*, Orthogonality of Separation in Two-Dimensional Liquid Chromatography, *Anal. Chem.*, **2005**, 77 (19), 6426-6434.
172. Delmotte, N., Lasaosa, M.*, et al.*, Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis, *J. Proteome Res.*, **2007**, 6 (11), 4363-4373.
173. Wang, Y., Yang, F.*, et al.*, Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells, *Proteomics*, **2011**, 11 (10), 2019-2026.
174. Batth, T. S., Francavilla, C.*, et al.*, Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics, *J. Proteome Res.*, **2014**, 13 (12), 6176-6186.
175. Taouatas, N., Altelaar, A. F. M.*, et al.*, Strong Cation Exchange-based Fractionation of Lys-N-generated Peptides Facilitates the Targeted Analysis of Post-translational Modifications, *Mol. Cell. Proteomics*, **2009**, 8 (1), 190-200.
176. Hennrich, M. L., Groenewold, V.*, et al.*, Improving depth in phosphoproteomics by using a strong cation exchange-weak anion exchange-reversed phase multidimensional separation approach, *Anal. Chem.*, **2011**, 83 (18), 7137-7143.
177. Dai, J., Jin, W.-H.*, et al.*, Protein Phosphorylation and Expression Profiling by Yin-Yang Multidimensional Liquid Chromatography (Yin-Yang MDLC) Mass Spectrometry, *J. Proteome Res.*, **2006**, 6 (1), 250-262.
178. Nie, S., Dai, J.*, et al.*, Comprehensive profiling of phosphopeptides based on anion exchange followed by flow-through enrichment with titanium dioxide (AFET), *J. Proteome Res.*, **2010**, 9 (9), 4585-4594.
179. Han, G., Ye, M.*, et al.*, Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography, *Proteomics*, **2008**, 8 (7), 1346-1361.
180. Alpert, A. J., Hudecz, O.*, et al.*, Anion-exchange chromatography of phosphopeptides: weak anion exchange versus strong anion exchange and anion-exchange chromatography versus electrostatic repulsion-hydrophilic interaction chromatography, *Anal. Chem.*, **2015**, 87 (9), 4704-4711.
181. Alpert, A. J., Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds, *J. Chromatogr. A*, **1990**, 499 (Supplement C), 177-196.
182. Alpert, A. J., Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides, *Anal. Chem.*, **2008**, 80 (1), 62-76.
183. Gan, C. S., Guo, T.*, et al.*, A comparative study of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) versus SCX-IMAC-based methods for phosphopeptide isolation/enrichment, *J. Proteome Res.*, **2008**, 7 (11), 4869-4877.
184. Hao, P., Guo, T.*, et al.*, Simultaneous analysis of proteome, phospho- and glycoproteome of rat kidney tissue with electrostatic repulsion hydrophilic interaction chromatography, *PloS one*, **2011**, 6 (2), e16884.
185. Chien, K. Y., Liu, H. C.*, et al.*, Development and application of a phosphoproteomic method using electrostatic repulsion-hydrophilic interaction chromatography (ERLIC), IMAC, and LC-MS/MS analysis to study Marek's Disease Virus infection, *J. Proteome Res.*, **2011**, 10 (9), 4041-4053.
186. Lanucara, F., Chi Hoo Lee, D.*, et al.*, Unblocking the Sink: Improved CID-Based Analysis of Phosphorylated Peptides by Enzymatic Removal of the Basic C-Terminal Residue, *J. Am. Soc. Mass Spectr.*, **2014**, 25 (2), 214-225.
187. Boersema, P. J., Mohammed, S.*, et al.*, Phosphopeptide fragmentation and analysis by mass spectrometry, *J. Mass Spectrom.*, **2009**, 44 (6), 861-878.

188. DeGnore, J. and Qin, J., Fragmentation of phosphopeptides in an ion trap mass spectrometer, *J. Am. Soc. Mass Spectr.*, **1998**, 9 (11), 1175-1188.
189. Gronert, S., Li, K. H.*, et al.*, Manipulating the fragmentation patterns of phosphopeptides via gas-phase boron derivatization: determining phosphorylation sites in peptides with multiple serines, *J. Am. Soc. Mass Spectrom.*, **2005**, 16 (12), 1905-1914.
190. Schroeder, M. J., Shabanowitz, J.*, et al.*, A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry, *Anal. Chem.*, **2004**, 76 (13), 3590-3598.
191. Villen, J., Beausoleil, S. A.*, et al.*, Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis, *Proteomics*, **2008**, 8 (21), 4444-4452.
192. Ulintz, P. J., Yocum, A. K.*, et al.*, Comparison of MS(2)-only, MSA, and MS(2)/MS(3) methodologies for phosphopeptide identification, *J. Proteome Res.*, **2009**, 8 (2), 887-899.
193. Nagaraj, N., D'Souza, R. C.*, et al.*, Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation, *J. Proteome Res.*, **2010**, 9 (12), 6786-6794.
194. Lundby, A., Secher, A.*, et al.*, Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues, *Nat. Commun.*, **2012**, 3, 876.
195. Boja, E. S., Phillips, D.*, et al.*, Quantitative Mitochondrial Phosphoproteomics Using iTRAQ on an LTQ-Orbitrap with High Energy Collision Dissociation, *J. Proteome Res.*, **2009**, 8 (10), 4665-4675.
196. Jedrychowski, M. P., Huttlin, E. L.*, et al.*, Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics, *Mol. Cell. Proteomics*, **2011**, 10 (12), M111.009910.
197. Iliuk, A., Jayasundera, K.*, et al.*, In-Depth Analyses of B Cell Signaling Through Tandem Mass Spectrometry of Phosphopeptides Enriched by PolyMAC, *International journal of mass spectrometry*, **2015**, 377, 744-753.
198. Kim, M. S., Zhong, J.*, et al.*, Systematic evaluation of alternating CID and ETD fragmentation for phosphorylated peptides, *Proteomics*, **2011**, 11 (12), 2568-2572.
199. Wang, L. D., Ficarro, S. B.*, et al.*, Phosphoproteomic profiling of mouse primary HSPCs reveals new regulators of HSPC mobilization, *Blood*, **2016**, 128 (11), 1465-1474.
200. Ferries, S., Perkins, S.*, et al.*, Evaluation of Parameters for Confident Phosphorylation Site Localization Using an Orbitrap Fusion Tribrid Mass Spectrometer, *J. Proteome Res.*, **2017**, 16 (9), 3448.
201. Perkins, D. N., Pappin, D. J.*, et al.*, Probability-based protein identification by searching sequence databases using mass spectrometry data, *Electrophoresis*, **1999**, 20 (18), 3551-3567.
202. Bradshaw, R. A., Burlingame, A. L.*, et al.*, Reporting protein identification data: the next generation of guidelines, *Mol. Cell. Proteomics*, **2006**, 5 (5), 787-788.
203. Elias, J. E. and 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.
204. Beausoleil, S. A., Villen, J.*, et al.*, A probability-based approach for high-throughput protein phosphorylation analysis and site localization, *Nat. Biotechnol.*, **2006**, 24 (10), 1285-1292.
205. Bailey, C. M., Sweet, S. M.*, et al.*, SLoMo: automated site localization of modifications from ETD/ECD mass spectra, *J. Proteome Res.*, **2009**, 8 (4), 1965-1971.

206. Taus, T., Köcher, T.*, et al.*, Universal and Confident Phosphorylation Site Localization Using phosphoRS, *J. Proteome Res.*, **2011**, 10 (12), 5354-5362.
207. Savitski, M. M., Lemeer, S.*, et al.*, Confident phosphorylation site localization using the Mascot Delta Score, *Mol. Cell. Proteomics*, **2011**, 10 (2), M110.003830.
208. Baker, P. R., Trinidad, J. C.*, et al.*, Modification site localization scoring integrated into a search engine, *Mol. Cell. Proteomics*, **2011**, 10 (7), M111.008078.
209. Chalkley, R. J. and Clauser, K. R., Modification site localization scoring: strategies and performance, *Mol. Cell. Proteomics*, **2012**, 11 (5), 3-14.
210. Attwood, P. V., Piggott, M. J.*, et al.*, Focus on phosphohistidine, *Amino Acids*, **2007**, 32 (1), 145-156.
211. Hultquis, D. E., Moyer, R. W.*, et al.*, Preparation and characterization of 1-phosphohistidine and 3-phosphohistidine, *Biochemistry*, **1966**, 5 (1), 322-&.
212. Hultquis, D. E., Preparation and characterisation of phosphorylated derivatives of histidine, *Biochim. Biophys. Acta.*, **1968**, 153 (2), 329-&.
213. Wei, Y. F. and Matthews, H. R., Identification of phopshohistidine in proteins and purification of protein-histidine kinases, *Methods Enzymol.*, **1991**, 200, 388-414.
214. Fujitaki, J. M. and Smith, R. A., Techniques in the detection and characterization of phosphoramidate-containing proteins, *Methods Enzymol.*, **1984**, 107, 23-36.
215. Napper, S., Kindrachuk, J.*, et al.*, Selective extraction and characterization of a histidine-phosphorylated peptide using immobilized copper(II) ion affinity chromatography and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Anal. Chem.*, **2003**, 75 (7), 1741-1747.
216. Kleinnijenhuis, A. J., Kjeldsen, F.*, et al.*, Analysis of histidine phosphorylation using tandem MS and ion - Electron reactions, *Anal. Chem.*, **2007**, 79 (19), 7450-7456.
217. Gonzalez-Sanchez, M. B., Lanucara, F.*, et al.*, Attempting to rewrite History: challenges with the analysis of histidine-phosphorylated peptides, *Biochem. Soc. T.*, **2013**, 41, 1089-1095.
218. Lapek, J. D., Tombline, G.*, et al.*, Evidence of histidine and aspartic acid phosphorylation in human prostate cancer cells, *N-S Arch. Pharmacol.*, **2015**, 388 (2), 161-173.
219. Medzihradszky, K. F., Phillipps, N. J.*, et al.*, Synthesis and characterization of histidine-phosphorylated peptides, *Protein Sci.*, **1997**, 6 (7), 1405-1411.
220. Schenkels, C., Erni, B.*, et al.*, Phosphofurylalanine, a stable analog of phosphohistidine, *Bioorg. Med. Chem. Lett.*, **1999**, 9 (10), 1443-1446.
221. Pirrung, M. C., James, K. D.*, et al.*, Thiophosphorylation of histidine, *J. Org. Chem.*, **2000**, 65 (25), 8448-8453.
222. McAllister, T. E., Nix, M. G.*, et al.*, Fmoc-chemistry of a stable phosphohistidine analogue, *Chem. Commun. (Camb)*, **2011**, 47 (4), 1297-1299.
223. Mukai, S., Flematti, G. R.*, et al.*, Stable triazolylphosphonate analogues of phosphohistidine, *Amino Acids*, **2012**, 43 (2), 857-874.
224. Kee, J. M., Villani, B.*, et al.*, Development of stable phosphohistidine analogues, *J. Am. Chem. Soc.*, **2010**, 132 (41), 14327-14329.
225. Kee, J.-M., Oslund, R. C.*, et al.*, A pan-specific antibody for direct detection of protein histidine phosphorylation, *Nat. Chem. Biol.*, **2013**, 9 (7), 416-U428.
226. Kee, J. M., Oslund, R. C.*, et al.*, A second-generation phosphohistidine analog for production of phosphohistidine antibodies, *Org. Lett.*, **2015**, 17 (2), 187-189.
227. Fuhs, S. R., Meisenhelder, J.*, et al.*, Monoclonal 1- and 3-Phosphohistidine Antibodies: New Tools to Study Histidine Phosphorylation, *Cell*, **2015**, 162 (1), 198-210.

228. Benjamini, Y. and Hochberg, Y., Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *Journal of the Royal Statistical Society. Series B (Methodological)*, **1995**, 57 (1), 289-300.
229. Gates, M. B., Tomer, K. B.*, et al.*, Comparison of metal and metal oxide media for phosphopeptide enrichment prior to mass spectrometric analyses, *J. Am. Soc. Mass Spectrom.*, **2010**, 21 (10), 1649-1659.
230. Attwood, P. V., Ludwig, K.*, et al.*, Chemical phosphorylation of histidine-containing peptides based on the sequence of histone H4 and their dephosphorylation by protein histidine phosphatase, *BBA-Proteins Proteom.*, **2010**, 1804 (1), 199-205.
231. Hohenester, U. M., Ludwig, K.*, et al.*, Chemical phosphorylation of histidine residues in proteins using potassium phosphoramidate -- a tool for the analysis of acid-labile phosphorylation, *Curr. Drug Deliv.*, **2013**, 10 (1), 58-63.
232. Pinkse, M. W., Mohammed, S.*, et al.*, Highly robust, automated, and sensitive online TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster, *J. Proteome Res.*, **2008**, 7 (2), 687-697.
233. Wilson-Grady, J. T., Villen, J.*, et al.*, Phosphoproteome analysis of fission yeast, *J. Proteome Res.*, **2008**, 7 (3), 1088-1097.
234. Schmidt, A., Trentini, D. B.*, et al.*, Quantitative phosphoproteomics reveals the role of protein arginine phosphorylation in the bacterial stress response, *Mol. Cell. Proteomics*, **2014**, 13 (2), 537-550.
235. Miles, E. W., [41] Modification of histidyl residues in proteins by diethylpyrocarbonate in *Methods in enzymology*, eds. Hirs, C. H. W. and Serge, N. T., Academic Press, **1977**, pp. 431-442.
236. Uchida, K. and Stadtman, E. R., Modification of histidine residues in proteins by reaction with 4-hydroxynonenal, *Proc. Natl. Acad. Sci. USA*, **1992**, 89 (10), 4544-4548.
237. Kowalewska, K., Stefanowicz, P.*, et al.*, Electron capture dissociation mass spectrometric analysis of lysine-phosphorylated peptides, *Bioscience Rep.*, **2010**, 30 (Pt 6), 433-443.
238. Ek, P., Ek, B.*, et al.*, Phosphohistidine phosphatase 1 (PHPT1) also dephosphorylates phospholysine of chemically phosphorylated histone H1 and polylysine, *Ups. J. Med. Sci.*, **2015**, 120 (1), 20-27.
239. Duclos, B., Marcandier, S.*, et al.*, [2] Chemical properties and separation of phosphoamino acids by thin-layer chromatography and/or electrophoresis, *Methods Enzymol.*, **1991**, 201, 10-21.
240. Krenkova, J., Lacher, N. A.*, et al.*, Control of Selectivity via Nanochemistry: Monolithic Capillary Column Containing Hydroxyapatite Nanoparticles for Separation of Proteins and Enrichment of Phosphopeptides, *Anal. Chem.*, **2010**, 82 (19), 8335-8341.
241. Mellacheruvu, D., Wright, Z.*, et al.*, The CRAPome: a contaminant repository for affinity purification-mass spectrometry data, *Nat. Methods*, **2013**, 10 (8), 730-736.
242. Rees, J. S., Lilley, K. S.*, et al.*, The chicken B-cell line DT40 proteome, beadome and interactomes, *Data in Brief*, **2015**, 3, 29-33.
243. Beausoleil, S. A., Jedrychowski, M.*, et al.*, Large-scale characterization of HeLa cell nuclear phosphoproteins, *Proc. Natl. Acad. Sci. USA*, **2004**, 101 (33), 12130-12135.
244. Dai, J., Wang, L. S.*, et al.*, Fully automatic separation and identification of phosphopeptides by continuous pH-gradient anion exchange online coupled with reversed-phase liquid chromatography mass spectrometry, *J. Proteome Res.*, **2009**, 8 (1), 133-141.

245. Ficarro, S. B., Zhang, Y.*, et al.*, Online nanoflow multidimensional fractionation for high efficiency phosphopeptide analysis, *Mol. Cell. Proteomics*, **2011**, 10 (11), O111.011064.
246. Pace, C. N., Grimsley, G. R.*, et al.*, Protein Ionizable Groups: pK Values and Their Contribution to Protein Stability and Solubility, *J. Biol. Chem.*, **2009**, 284 (20), 13285-13289.
247. Tian, M., Cheng, H.*, et al.*, Phosphoproteomic Analysis of the Highly-Metastatic Hepatocellular Carcinoma Cell Line, MHCC97-H, *Int. J. Mol. Sci.*, **2015**, 16 (2), 4209.
248. Rappsilber, J., Mann, M.*, et al.*, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, *Nat. Protocols*, **2007**, 2 (8), 1896-1906.
249. Ishihama, Y., Rappsilber, J.*, et al.*, Modular Stop and Go Extraction Tips with Stacked Disks for Parallel and Multidimensional Peptide Fractionation in Proteomics, *J. Proteome Res.*, **2006**, 5 (4), 988-994.
250. Boersema, P. J., Foong, L. Y.*, et al.*, In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling, *Mol. Cell. Proteomics*, **2010**, 9 (1), 84-99.
251. Sharma, K., D'Souza, Rochelle C. J.*, et al.*, Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling, *Cell Reports*, **2014**, 8 (5), 1583-1594.
252. Oslund, R. C., Kee, J.-M.*, et al.*, A Phosphohistidine Proteomics Strategy Based on Elucidation of a Unique Gas-Phase Phosphopeptide Fragmentation Mechanism, *J. Am. Chem. Soc.*, **2014**, 136 (37), 12899-12911.
253. http://www.proteinatlas.org/search/PHPT1, The Human Protein Atlas, (accessed Sept, 2017).
254. Xu, A., Hao, J.*, et al.*, 14-kDa phosphohistidine phosphatase and its role in human lung cancer cell migration and invasion, *Lung cancer (Amsterdam, Netherlands)*, **2010**, 67 (1), 48-56.
255. Shen, H., Yang, P.*, et al.*, Nuclear expression and clinical significance of phosphohistidine phosphatase 1 in clear-cell renal cell carcinoma, *J. Int. Med. Res.*, **2015**, 43 (6), 747-757.
256. Scholten, A., Preisinger, C.*, et al.*, Phosphoproteomics Study Based on In Vivo Inhibition Reveals Sites of Calmodulin-Dependent Protein Kinase II Regulation in the Heart, *J. Am. Heart Assoc.*, **2013**, 2 (4).
257. Mi, H., Muruganujan, A.*, et al.*, Large-scale gene function analysis with the PANTHER classification system, *Nat. Protocols*, **2013**, 8 (8), 1551-1566.
258. Zhang, Y., Ficarro, S. B.*, et al.*, Optimized Orbitrap HCD for quantitative analysis of phosphopeptides, *J. Am. Soc. Mass Spectrom.*, **2009**, 20 (8), 1425-1434.
259. Dennis, G., Jr., Sherman, B. T.*, et al.*, DAVID: Database for Annotation, Visualization, and Integrated Discovery, *Genome Biol.*, **2003**, 4 (5), P3.
260. Huang, D. W., Sherman, B. T.*, et al.*, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat. Protocols*, **2008**, 4 (1), 44-57.
261. Khatri, P. and Drăghici, S., Ontological analysis of gene expression data: current tools, limitations, and open problems, *Bioinformatics*, **2005**, 21 (18), 3587-3595.
262. Schwartz, D. and Gygi, S. P., An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets, *Nat. Biotechnol.*, **2005**, 23 (11), 1391-1398.
263. Chou, M. F. and Schwartz, D., Biological sequence motif discovery using motif-x, *Curr. Protoc. Bioinformatics*, **2011**, Chapter 13, Unit 13.15-24.

264. Zhang, L., Wang, J. C*., et al.*, Functional Role of Histidine in the Conserved His-x-Asp Motif in the Catalytic Core of Protein Kinases, *Sci. Rep.*, **2015**, 5, 10115.
265. Verba, K. A., Wang, R. Y*., et al.*, Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase, *Science*, **2016**, 352 (6293), 1542-1547.
266. Verba, K. A. and Agard, D. A., How Hsp90 and Cdc37 Lubricate Kinase Molecular Switches, *Trends Biochem. Sci.*, **2017**.
267. Bi, W., Xiao, L*., et al.*, c-Jun N-terminal kinase enhances MST1-mediated pro-apoptotic signaling through phosphorylation at serine 82, *J. Biol. Chem.*, **2010**, 285 (9), 6259-6264.
268. Liu, W., Wu, J*., et al.*, Regulation of neuronal cell death by c-Abl-Hippo/MST2 signaling pathway, *PloS one*, **2012**, 7 (5), e36562.
269. Zetterqvist, O. and Engstrom, L., Isolation of N-e-[32P]phosphoryl-lysine from rat-liver cell sap after incubation with [32P]adenosine triphosphate, *Biochim. Biophys. Acta*, **1967**, 141 (3), 523-532.
270. Ohmori, H., Kuba, M*., et al.*, Two phosphatases for 6-phospholysine and 3-phosphohistidine from rat brain, *J. Biol. Chem.*, **1993**, 268 (11), 7625-7627.
271. Wong, C., Faiola, B*., et al.*, Phosphohistidine and phospholysine phosphatase activities in the rat: potential protein-lysine and protein-histidine phosphatases?, *Biochem. J.*, **1993**, 296 (Pt 2), 293-296.
272. Ek, P., Ek, B*., et al.*, Phosphohistidine phosphatase 1 (PHPT1) also dephosphorylates phospholysine of chemically phosphorylated histone H1 and polylysine, *Ups. J. Med. Sci.*, **2015**, 120 (1), 20-27.
273. Fuhrmann, J., Schmidt, A*., et al.*, McsB is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator CtsR, *Science*, **2009**, 324 (5932), 1323-1327.
274. Wakim, B. T. and Aswad, G. D., Ca(2+)-calmodulin-dependent phosphorylation of arginine in histone 3 by a nuclear kinase from mouse leukemia cells, *J. Biol. Chem.*, **1994**, 269 (4), 2722-2727.
275. Levy-Favatier, F., Delpech, M*., et al.*, Characterization of an arginine-specific protein kinase tightly bound to rat liver DNA, *Eur. J. Biochem.*, **1987**, 166 (3), 617-621.
276. Smith, L. S., Kern, C. W*., et al.*, Phosphorylation on basic amino acids in myelin basic protein, *Biochem. Bioph. Res. Co.*, **1976**, 71 (2), 459-465.
277. Nishino, M., Tsujimura, S*., et al.*, N omega-phosphoarginine phosphatase from rat renal microsome was alkaline phosphatase, *Arch. Biochem. Biophys.*, **1994**, 312 (1), 101-106.
278. Kumon, A., Kodama, H*., et al.*, N(omega)-phosphoarginine phosphatase (17 kDa) and alkaline phosphatase as protein arginine phosphatases, *J. Biochem.*, **1996**, 119 (4), 719-724.
279. Ridder, I. S. and Dijkstra, B. W., Identification of the Mg2+-binding site in the P-type ATPase and phosphatase members of the HAD (haloacid dehalogenase) superfamily by structural similarity to the response regulator protein CheY, *Biochem. J.*, **1999**, 339 (Pt 2), 223-226.
280. Collet, J. F., Stroobant, V*., et al.*, Mechanistic studies of phosphoserine phosphatase, an enzyme related to P-type ATPases, *J. Biol. Chem.*, **1999**, 274 (48), 33985-33990.
281. Hu, C. A., Delauney, A. J*., et al.*, A bifunctional enzyme (delta 1-pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants, *Proc. Natl. Acad. Sci. USA*, **1992**, 89 (19), 9354-9358.

282. Aral, B., Schlenzig, J. S*., et al.*, Database cloning human delta 1-pyrroline-5-carboxylate synthetase (P5CS) cDNA: a bifunctional enzyme catalyzing the first 2 steps in proline biosynthesis, *Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie*, **1996**, 319 (3), 171-178.
283. Griffith, O. W. and Mulcahy, R. T., The enzymes of glutathione synthesis: gamma-glutamylcysteine synthetase, *Advances in enzymology and related areas of molecular biology*, **1999**, 73, 209-267, xii.
284. Wang, R. H., Tao, L*., et al.*, Turnover of the acyl phosphates of human and murine prothymosin alpha in vivo, *J. Biol. Chem.*, **1997**, 272 (42), 26405-26412.
285. Tao, L., Wang, R. H*., et al.*, Metabolic regulation of protein-bound glutamyl phosphates: insights into the function of prothymosin alpha, *J. Cell. Physiol.*, **1999**, 178 (2), 154-163.
286. Cohen-Solal, L., Cohen-Solal, M*., et al.*, Identification of gamma-glutamyl phosphate in the alpha 2 chains of chicken bone collagen, *Proc. Natl. Acad. Sci. USA*, **1979**, 76 (9), 4327-4330.
287. Peterson, C. L. and Laniel, M. A., Histones and histone modifications, *Current biology : CB*, **2004**, 14 (14), R546-551.
288. Tanner, S., Shu, H*., et al.*, InsPecT: identification of posttranslationally modified peptides from tandem mass spectra, *Anal. Chem.*, **2005**, 77 (14), 4626-4639.
289. Savitski, M. M., Nielsen, M. L*., et al.*, ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures, *Mol. Cell. Proteomics*, **2006**, 5 (5), 935-948.
290. Han, X., He, L*., et al.*, PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications, *J. Proteome Res.*, **2011**, 10 (7), 2930-2936.

Chapter 9. Appendix

The following excel files/tables are provided in electronic format:

Table 9.1 Protein IDs shared with Fuhs *et al.*

Proteins identified by LC-MS/MS following immunoprecipitation (IP) with 1- and 3-pHis antibodies (2 or more peptides required for protein ID), that were previously identified by a similar IP experiment. Proteins from the Fuhs *et al.* list were filtered to only include those which showed a greater than 2-fold difference compared to the control. 317 (28%) and 295 (24%) proteins were shared between this analysis and the previously published data for the 1- and 3-pHis IP experiments respectively.

Table 9.2 pHis_1% FLR

Table 9.3 pLys_1% FLR

Table 9.4 pArg_1% FLR

Table 9.5 pAsp_1% FLR

Table 9.6 pGlu_1% FLR

List of pX peptides/proteins identified in each of the two conditions (PHPT1 siRNA and NT siRNA). Where a peptide was identified more than once the highest scoring instance is shown, with a column to indicate those peptides with more than one observation. The 'pX site' column indicates the site(s) that are above the 1% FLR score cut-off, with all other phosphosites (below the score cut-off and/or not the residue of interest) shown in a separate column. For pHis peptides, the triplet score is shown.

Table 9.7 DAVID_pHis

Table 9.8 DAVID_pLys

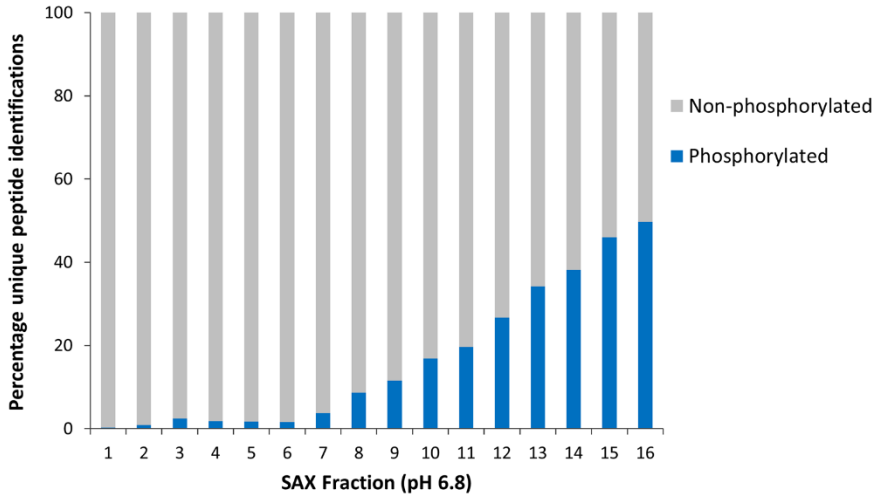
Table 9.9 DAVID_pArg

Table 9.10 DAVID_pAsp

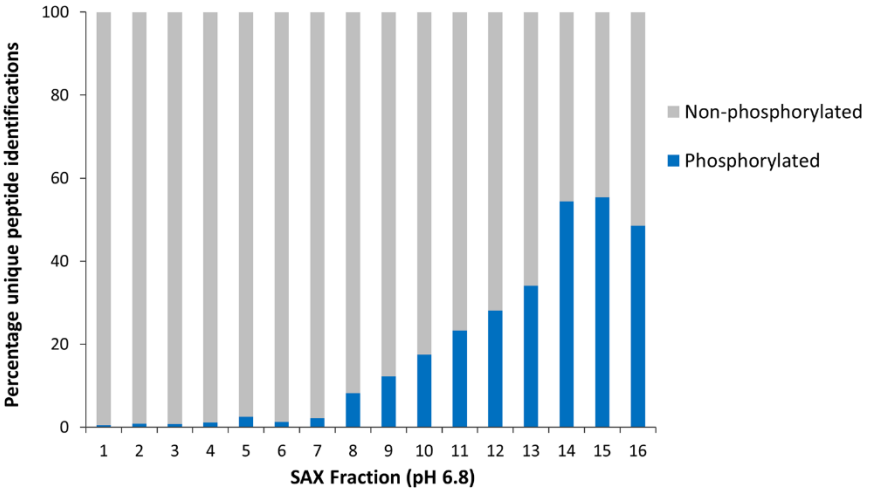
Table 9.11 DAVID_pGlu

Output from functional annotation clustering using DAVID (version 6.8). List of pX proteins identified in each condition (NT siRNA and PHPT1 siRNA) compared to background of all proteins identified by analysis of these samples. Classification stringency: medium; similarity term overlap: 3; initial/final group membership: 3; EASE: 1.0

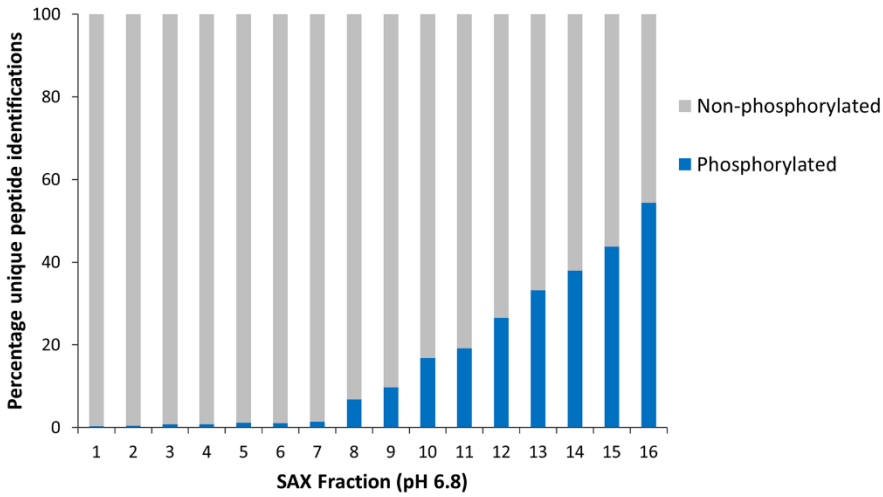
a


Percentage unique peptide identifications
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
SAX Fraction (pH 6.8)
Non-phosphorylated
Phosphorylated

b


Percentage unique peptide identifications
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
SAX Fraction (pH 6.8)
Non-phosphorylated
Phosphorylated

c


Percentage unique peptide identifications
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
SAX Fraction (pH 6.8)
Non-phosphorylated
Phosphorylated

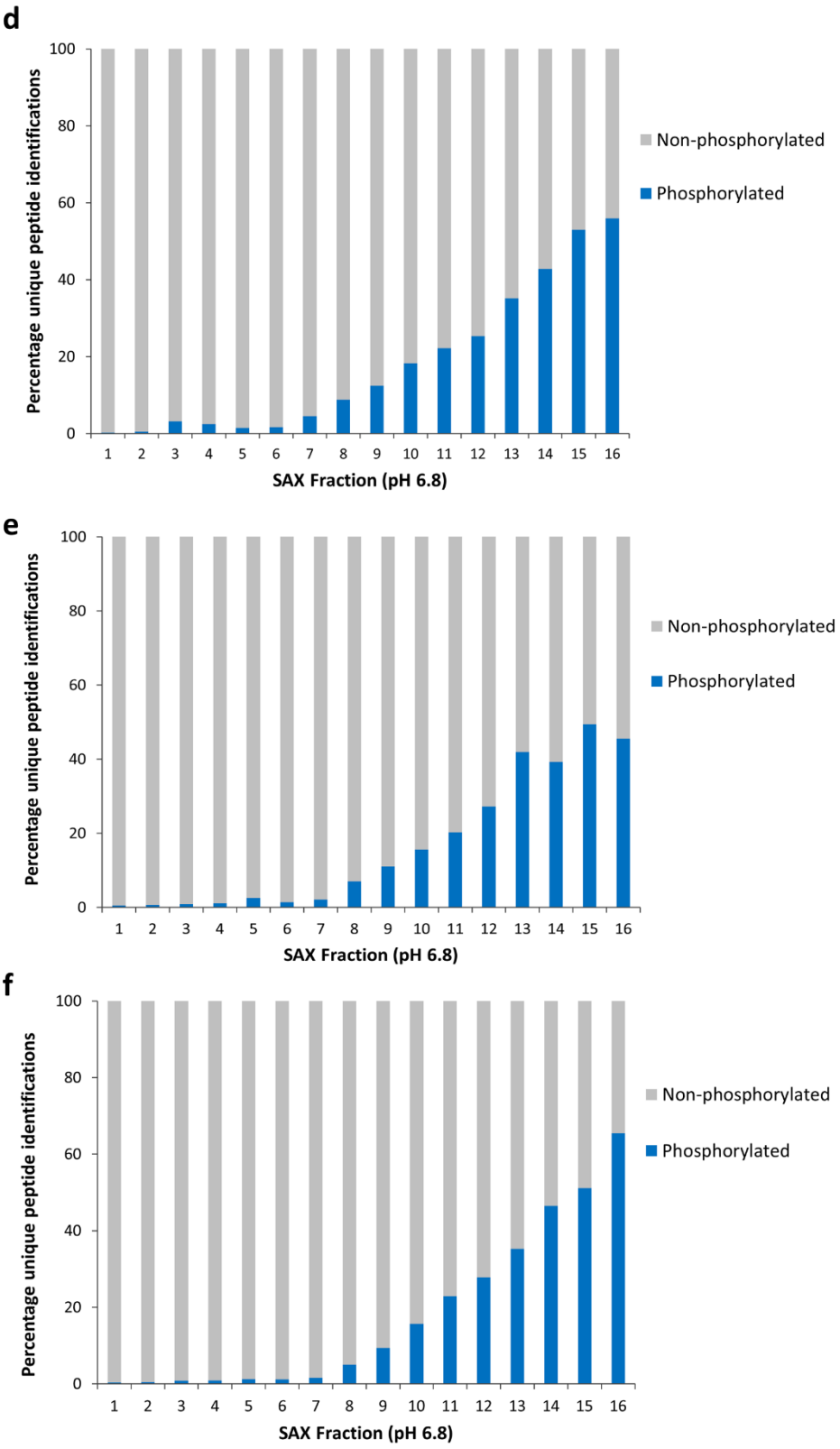
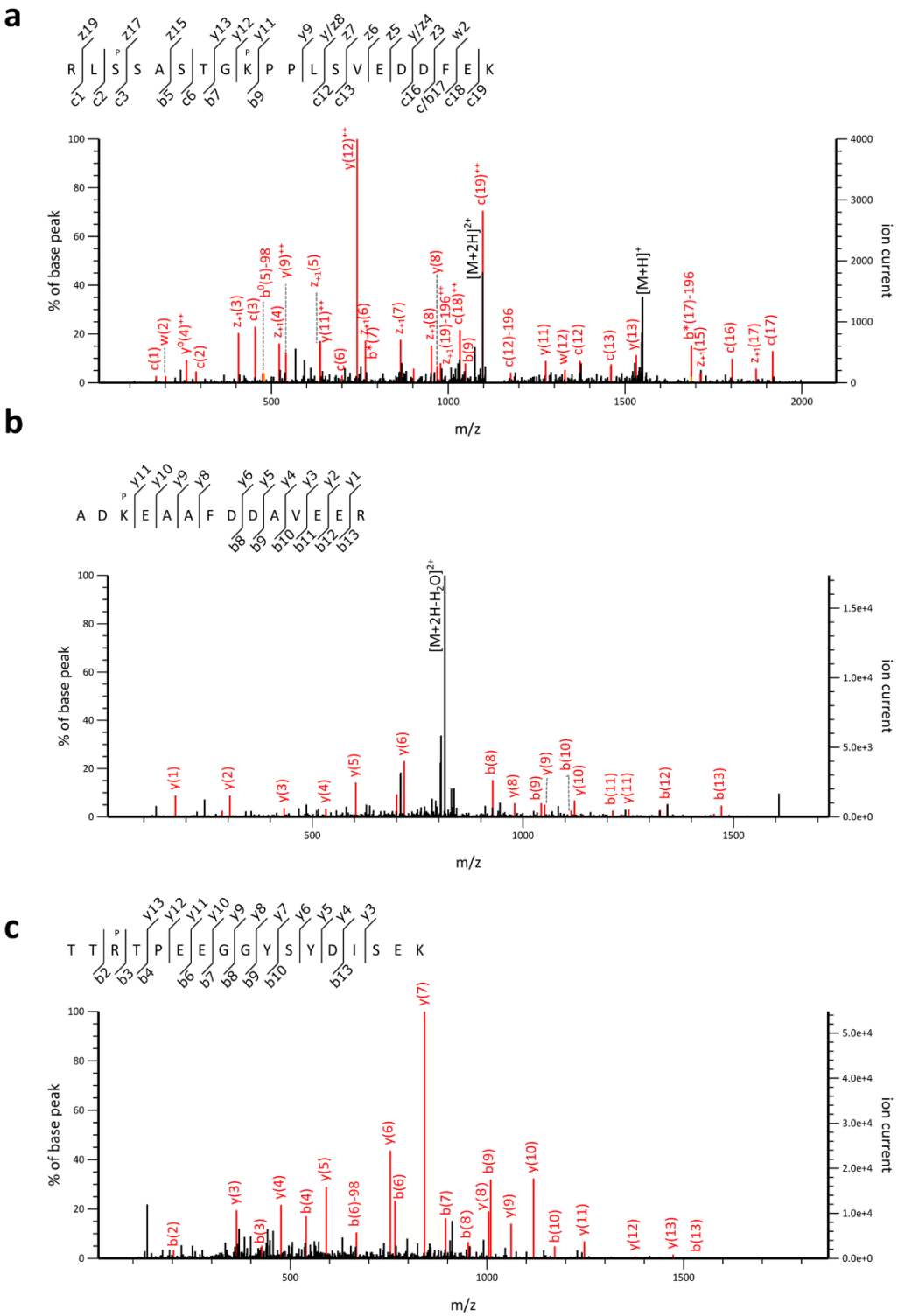
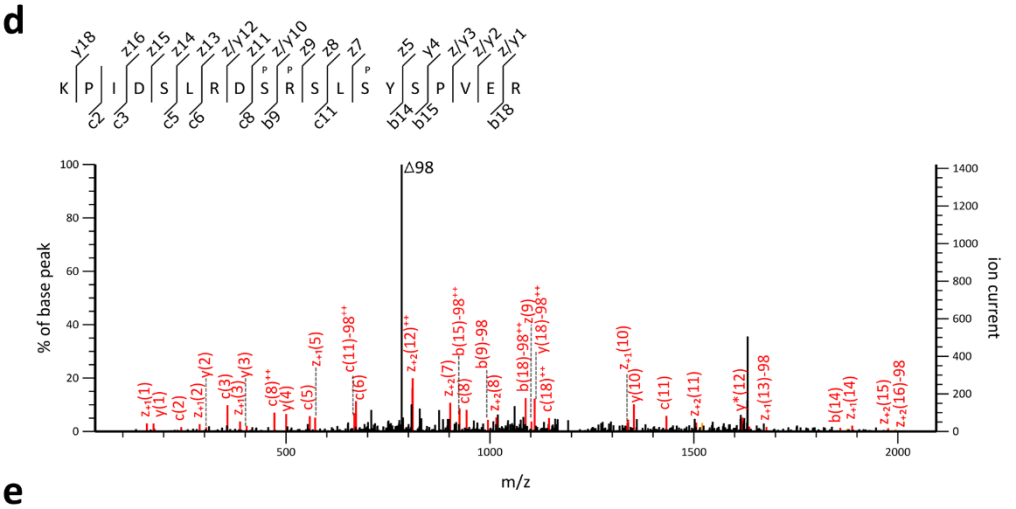


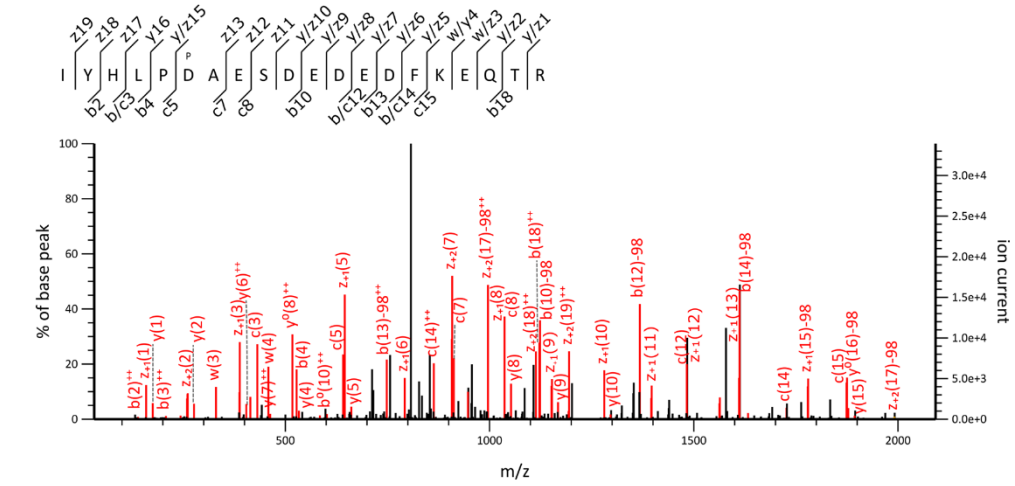
Figure 9.1 Percentage of unique phosphorylated and non-phosphorylated peptides in each fraction following SAX fractionation of HeLa lysate at pH 6.8, for all six bioreplicates **a)** PHPT1 siRNA biorep 1, **b)** PHPT1 siRNA biorep 2, **c)** PHPT1 siRNA biorep 3, **d)** NT siRNA biorep 1 (also shown in main part of thesis: *Chapter 5. Results III*, *Figure 5.4*), **e)** NT siRNA biorep 2 and **f)** NT siRNA biorep


a
R L S(p) S A S T G K(p) P P L S V E D D F E K
% of base peak
ion current
m/z
b
A D K(p) E A A F D D A V E E R
[M+2H-H2O]2+
% of base peak
ion current
m/z
c
T T R(p) T P E E G G Y S Y D I S E K
% of base peak
ion current
m/z

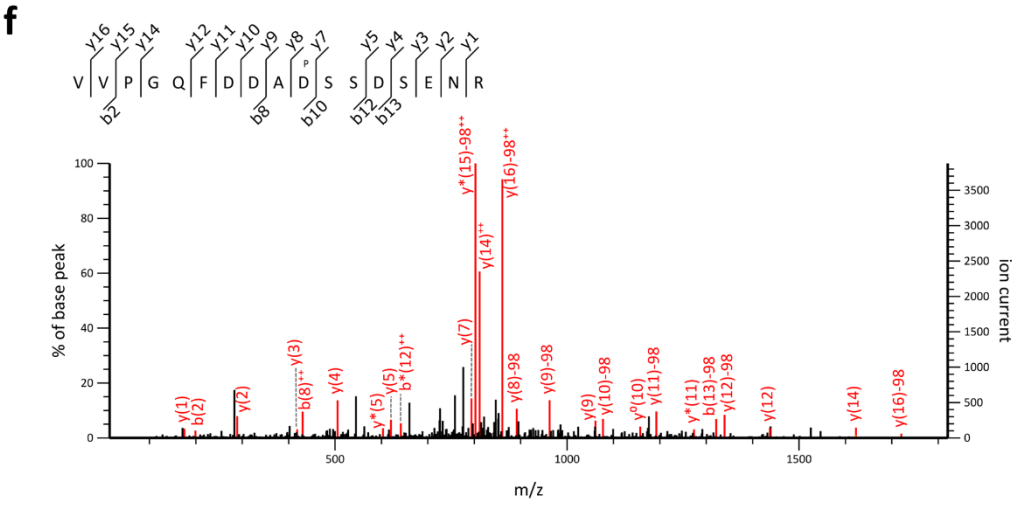
d


K P I D S L R D S S R S L S Y S P V E R
% of base peak
ion current
m/z
Δ98

e


I Y H L P D A E S D E D E D F K E Q T R
% of base peak
ion current
m/z

f


V V P G Q F D D A D S S D S E N R
% of base peak
ion current
m/z

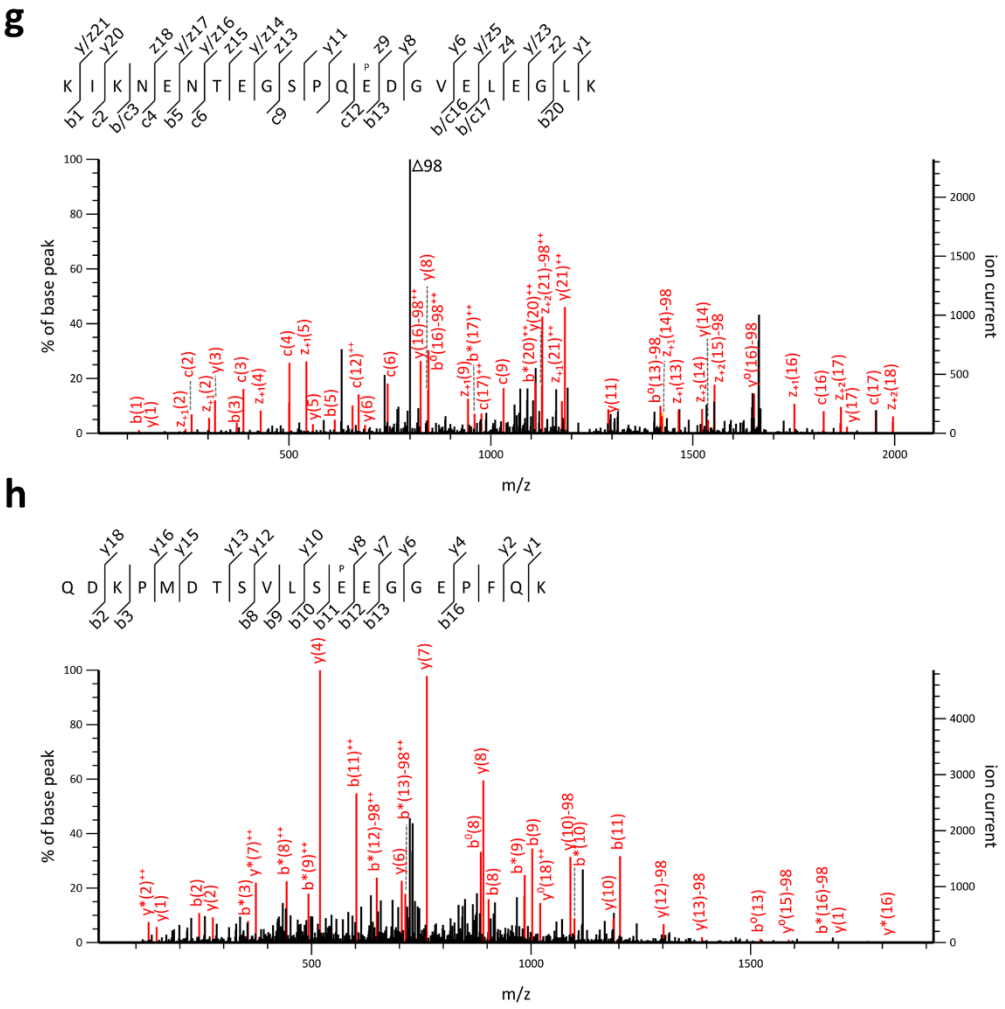


Figure 9.2 Example tandem MS spectra for each site of phosphorylation, following either HCD or EThcD fragmentation. **a)** pLys peptide: RLpSSASTGpKPPLSVEDDFEK, *m/z* 775.01, EThcD, ptmRS scores S3: 99.1, K9: 96.37; **b)** pLys peptide: ADpKEAAFDDAVEER, *m/z* 823.33, HCD, ptmRS score K3: 100; **c)** pArg peptide: TTpRTPEEGGYSYDISEK, *m/z* 671.62, HCD, ptmRS score R3: 99.93; **d)** pArg peptide: KPIDSLRDpSpRSLpSYSPVER, *m/z* 815.69, EThcD, ptmRS scores S9: 100, R10: 99.44, S13: 84.84; **e)** pAsp peptide: IYHLPpDAESDEDEDFKEQTR, *m/z* 839.69, EThcD, ptmRS score D6: 99.87; **f)** pAsp peptide: VVPGQFDDApDSSDSENR, *m/z* 959.37, HCD, ptmRS score D10: 99.66; **g)** pGlu peptide: KIKNENTEGSPQpEDGVELEGLK, *m/z* 832.06, EThcD, ptmRS score E13: 97.93; **h)** pGlu peptide: QDKPMDTSVLSpEEGGEPFQK, *m/z* 768.00, HCD, ptmRS score E12: 99.37. Phosphorylated residue(s) are underlined.