Universität für Bodenkultur Wien
University of Natural Resources
and Life Sciences, Vienna

Master Thesis

Improving the Enrichment of Phosphorylated Tyrosine for Comprehensive Proteomic Profiling of Biopharmaceutical Production Cell Lines

Submitted by

Markus RIEDL, BSc

in the framework of the Master programme

Biotechnology

in partial fulfilment of the requirements for the academic degree

Diplom-Ingenieur

Vienna, August 2021

Supervisor:

Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Nicole Borth
Institute of Animal Cell Technology and Systems Biology
Department of Biotechnology

Affidavit

I hereby declare that I have authored this master thesis independently, and that I have not used any assistance other than that which is permitted. The work contained herein is my own except where explicitly stated otherwise. All ideas taken in wording or in basic content from unpublished sources or from published literature are duly identified and cited, and the precise references included.

I further declare that this master thesis has not been submitted, in whole or in part, in the same or a similar form, to any other educational institution as part of the requirements for an academic degree.

I hereby confirm that I am familiar with the standards of Scientific Integrity and with the guidelines of Good Scientific Practice, and that this work fully complies with these standards and guidelines.

Vienna, September 2021 Markus RIEDL (*manu propria*)

Acknowledgement

First and foremost, I want to thank my supervisor Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Nicole Borth for giving me the possibility to carry out my master thesis in Ireland.

I want to thank Dr. Paula Meleady for providing scientific guidance throughout my stay at the NICB in Dublin. I would also like to thank my colleagues in Dublin, particularly Dr. Orla Coleman, Dr. Prashant Kaushik and Dr. Michael Henry, for their great help and assistance during my work.

Thanks to my girlfriend Eva Welsch for her continued support and for joining me on this stay abroad.

Finally, I want to thank my family for their unrelenting support and facilitation of my education.

Contents

List of Figures

List of Tables

Abbreviations

AUC	Area Under the Curve
CHO	Chinese Hamster Ovary
CID	Collision Induced Dissociation
ECD	Electron Capture Dissociation
ESI	Electronspray Ionization
ETD	Electron Transfer Dissociation
HCD	Higher Energy C-Trap Dissociation
IMAC	Immobilized Metal Affinity Chromatography
IP	Immunoprecipitation
iTRAQ	Isobaric Tags for Relative and Absolute Quantitation
MALDI	Matrix-Assisted Laser Desorption Ionization
MOAC	Metal Oxide Affinity Chromatography
MSA	Multistage Activation
pS, pT, pY	Phosphoserine, Phosphothreonine, Phosphotyrosine
PSM	Peptide Spectrum Match
PTM	Post-translational Modification
SILAC	Stable Isotope Labelling by Amino acids in Cell culture
TMT	Tandem Mass Tag
TOF	Time-of-Flight

Abstract

The Chinese Hamster Ovary cell line is the most widely used cell factory for biopharmaceutical products and is steadily developed with the objective to improve product yields. Rational approaches to such engineering measures require knowledge of fundamental biological processes that ultimately affect the cell's properties such as cell growth, differentiation, apoptosis and signal transduction. The post-translational modification of tyrosine residues by phosphorylation plays a pivotal role in modulating many such essential cell processes and properties. However, the extreme sub-stoichiometric and transient nature of tyrosine phosphorylation makes proteomic analyses of this modification a challenging endeavour that can only be achieved by specific enrichment of peptides containing phosphotyrosine. This thesis aims to develop and optimize a procedure to enable qualitative and quantitative profiling of tyrosine phosphorylation of proteins involved in industrially relevant phenotypes. The implications of several aspects of the enrichment procedure were evaluated and compared both in terms of the number of the identified tyrosine-phosphorylated proteins as well as the confidence of phosphorylation site inference. The identified proteins were discussed and put into biological context by analysis of Gene Ontology terms. This work provides an optimized protocol that achieves the biologically most relevant information about phosphorylated tyrosine and gives a perspective on improving the detection of tyrosine phosphorylation.

Zusammenfassung

Die Chinese Hamster Ovary Zelllinie ist die bedeutendste Zellfabrik für biopharmazeutische Produkte und wird laufend weiterentwickelt mit dem Ziel Produkterträge zu erhöhen. Rationale Ansätze zu diesen Entwicklungsmaßnahmen fordern ein umfassendes Wissen über fundamentale biologische Prozesse welche zellulären Eigenschaften wie Zellwachstum, Differenzierung, Apoptose und Signalverarbeitung zugrunde liegen. Die posttranslationale Modifikation von Tyrosin-Seitenketten durch Phosphorylierung spielt eine bedeutende Rolle in der Modulation dieser essenziellen Zellprozesse und -eigenschaften. Die extreme substöchiometrische und flüchtige Natur der Tyrosin-Phosphorylierung macht proteomische Analysen dieser Modifikation zu einem herausfordernden Unterfangen welches bisher nur durch spezifische Anreicherung von Peptiden mit Phosphotyrosin gelingt. Ziel der vorliegenden Masterarbeit ist die Entwicklung und Optimierung eines Verfahrens zur qualitativen und quantitativen Bestimmung von Tyrosin-Phosphorylierung industriell relevanter Proteine. Die Bedeutung verschiedener Einflussfaktoren des Anreicherungsverfahrens wurde evaluiert und vergleichend analysiert, sowohl in Bezug auf die Anzahl der identifizierten phosphorylierten Tyrosin-Seitenketten, also auch in Bezug auf die Konfidenz der inferierten Phosphorylierung. Die somit identifizierten Proteine wurden erörtert sowie in biologischem Kontext gesetzt durch Analyse der Gen-Ontologie. Das erarbeitete Protokoll erzielt den biologisch bedeutsamsten Informationsgehalt zu phosphoryliertem Tyrosin und verdeutlicht Perspektiven zur Verbesserung der Detektion von Tyrosin-Phosphorylierung.

1 Introduction

1.1 Post-translational modifications and protein phosphorylation

Post-translational modifications (PTMs) are reversible chemical modifications of a protein's side-chain that add another layer of diversity and possible functionality. Alongside alternative splicing of RNA, PTMs ultimately extend the proteome of an organism significantly. More than 200 different types of these modifications are described [1] with some of the most known being additions of phosphate, glycosyl or methyl groups to amino acid residues.

While some PTMs are predominantly constant, like glycosylation, lipidation and disulifide bridges, others, like phosphorylation or ubiquitination, are more transient and subject to dynamic processes. The former are often important for correct protein folding and stability while the latter are involved in intracellular signaling or protein degradation [1].

PTMs have huge implications in the biotechnology industry. They are one of the main decision-making factors in the choice of the expression system for a biotechnological product. Despite their ease of cultivation and high growth rates, bacteria and yeast expression systems often fall short in terms of the required quality of products produced by them. More sophisticated compounds, especially if they are intended to be used as therapeutic proteins in human medicine, require human-like PTMs that often only mammalian cell factories can provide [2, 3]. Another way in which PTMs influence biotechnological processes is by regulating essential cell functions and properties like growth and signal transduction which ultimately affect product yields.

1.1.1 Phosphorylation

Phosphorylation of serine, threonine and tyrosine residues is by far the most abundant PTM [4, 5]. Protein kinases, enzymes dedciated to phosphorylation, are amongst the biggest protein superfamilies with over 500 members in both the human and mouse kinome for example [6, 7]. Phosphorylation is a highly dynamic PTM that drives the intracellular signal transduction cascades which connects cell surface receptors to

resulting phenotypes [8].

Phosphorylation regulates essential processes within eukaryotic cells. The most frequently reported process which is regulated by this PTM is transcription, followed by cell growth, cell cycle control and apoptosis. Functionally, phosphorylation most frequently modulates the molecular association of proteins, the activity of an enzyme or the intracellular localization [9].

The frequency of the three types of eukaryotic phosphorylation are not uniformly distributed. In vertebrates, tyrosine phosphorylation accounts for only about 0.05 % of all phosphorylations; phosphoserine $\sim$ 90 % and phosphothreonine $\sim$ 10 % [10].

1.1.2 Phosphotyrosine

Despite the occurrence of phosphorylated tyrosine residues in substochiometric amounts compared to other phosphorylations, the fraction of tyrosine kinases is much higher, representing 10–15 % of all protein kinases [11]. The low abundance also accounts for the late discovery of phosphorylation of tyrosine residues around 40 years ago, the last phosphorylation to be discovered [12, 13]. The chemical structure of phosphotyrosine can be seen in Figure 1.

Several reasons are described for this low abundance. First, tyrosine kinases are very tightly regulated and only activated in specific circumstances. Second, phosphotyrosines are very short-lived due to the high activity of phosphotyrosine phosphatases [14]. That is, unless they are protected, they are bound to a phosphotyrosine-binding (PTB) or Src-homology 2 (SH2) domain.

One major example of implications of tyrosine phosphorylation in intracellular processes are Receptor Tyrosine Kinases (RTKs), a large family of tyrosine kinases involved in cell signaling cascades. Almost a third of the whole known tyrosine kinome in humans are RTKs [15]. 20 subfamilies, of which mostly the extracellular receptor domains differ, are known. The intracellular tyrosine kinase domains are highly conserved throughout evolution which emphasises their essential role in cell signaling [16].

Apart from transmembrane RTKs, many non-receptor tyrosine kinases are involved in signaling as well. The well described kinase families of Src, Jak and Fak are associated

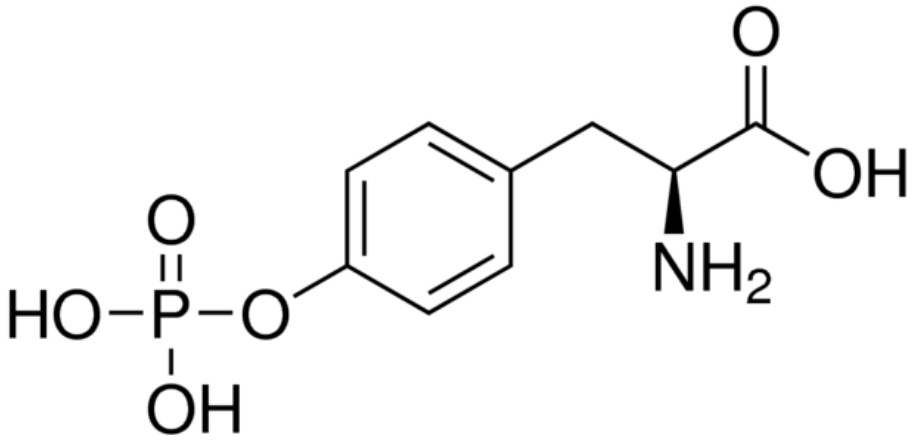

Figure 1: Chemical structure of phosphorylated tyrosine.

with membrane-spanning receptors which, by themselves, lack the ability to pass the signal on [10].

A thorough analysis of a cell's phosphotyrosine proteome can give extensive insight into fundamental cell processes that ultimately help to better understand and improve engineering measures.

1.2 Chinese Hamster Ovary cells

The biopharmaceutical industry requires cell factories that produce safe and correctly functional products. Due to their human-like PTMs, which assure the aforementioned criteria, mammalian cell lines are used for the production of most recombinant protein therapeutics. Chinese Hamster Ovary (CHO) cells are the primary work horse in this category with over 70 % of therapeutic proteins in the global market produced by them [17].

Initially, CHO cells were taken from a biopsy specimen of an inbred female Chinese hamster in 1957 for studies of their karyotype [18]. Their extensive use for the production of recombinant proteins began in the early 1980s with the first stable transfection of mammalian cells. Dihydrofolate reductase (DHFR) deficient strains of CHO cells were amongst the first to be transfected with these novel methods [19, 20, 21]. The first recombinant therapeutic protein approved for medical use in 1987 was produced by CHO cells [22].

The deficiency in DHFR activity enables the stable transfection with an expression vector that carries the gene of interest (GOI) as well as the necessary *dhfr* gene. Transfected cells can be grown under selective conditions. Subsequently, several resulting candidate cell lines are evaluated and subjected to further selective processes to obtain the most favorable phenotype in terms of growth and productivity. Alternatively to DHFR, glutamine synthetase (GS) can also be used as a selection marker in this process [22]. Furthermore, the addition of methotrexate (MTX) or methionine sulphoximide (MSX) to selection media for blocking DHFR or GS activity, respectively, increases the selective pressure towards strains with higher productivity [23].

1.2.1 Cell engineering and bioprocessing

Large-scale single cell suspension cultures in stirred-tank bioreactors, adapted from the cultivation of bacteria, were the first form of bioprocessing with these cell factories [24]. A variety of engineering measures to emphasize advantageous properties of cell lines as well as improvements in bioprocessing have led to drastic increases of productivity. While in the late 1980s, a batch run lasted for 7 days with a maximum cell density of 1–2 million cells ml^{-1} and a volumetric productivity of 50–100 $\mathrm{mg\,L}^{-1}$, todays fed-batch processes last 21 days with a density of 10–15 million cells ml^{-1} and a volumetric productivity of 1–5 $\mathrm{g\,L}^{-1}$ [25].

Means by which these advances were achieved include cellular engineering measures that adress apoptosis and autophagy [26, 27, 28] or lactate production [29, 30]. A variety of mechanisms can be employed to achieve engineering goals, ranging from the silencing and over- or underexpression of individual genes [31] to RNA-mediated techniques that can alter the expression of whole groups of genes using microRNAs (miRNA) [32, 33, 34].

Optimization in media composition and bioprocess conditions pose the second important field of improvement. Biphasic cultivation methods, for instance, have proved to be beneficial in many cases. Shifts in temperature and pH often lead to a reduced growth rate, or even arrest, and increased specific productivity, yielding more product overall [35, 36].

1.3 Proteomics

A comprehensive understanding and monitoring of cellular processes and developments is crucial for informed decisions and evaluations in engineering. This understanding comes best at multiple levels, from the genetic and transcriptional level up to the level of functional proteins that define the phenotype of a cell.

Proteomics data is used in many fields of current research. Advancements in sequencing technologies have led to an unparalleled expansion of genomic data. The task to functionally annotate these genomes, however, remains a major struggle that cannot be tackled by the genomic data alone. Using evidence-driven approaches, for example with the integration of data from protein studies, enables a reliable gene annotation and gene prediction [37].

In the context of the biopharmaceutical production of therapeutic proteins, it is very commonly of interest which molecular changes are triggered by certain measures and interventions. Differential proteomic studies shine light into the changes of a cell's proteome. A shift in temperature for biphasic cultivation of CHO cells, is one example of such differential proteomic studies. Kumar et al. found differential expression of proteins involved in apoptosis, cell size, proliferation, DNA repair and DNA replication in biphasic cultivated CHO cell cultures upon temperature shift [38]. In a subsequent study, Henry et al. looked specifically at the phosphoproteome of temperature-shifted cultures. Differential phosphoproteomics in biphasic cultivation revealed an even deeper understanding of biological changes at molecular level with biological processes like chromosome organization, cell cycle control and cell division down– and RNA processing upregulated at 31 °C [39].

1.3.1 Phosphoproteomics

Protein phosphorylation plays an essential role in signal transduction, apoptosis, cell proliferation and metabolism [40]. Therefore, phosphoproteomics has emerged to specifically study these important PTMs. A key step for every phosphoproteomic analysis is the enrichment of phophopeptides due to their low abundance overall. Phosphorylated serine, threonine and tyrosine residues are commonly enriched using either specific

antibodies or chemical resins.

The covalent incorporation of a phosphate group leads to a negative charge on the surface of the residue. This property is used for affinity enrichment of phosphorylated peptides with positively charged metal ions, like Fe^{3+}, Ga^{2+}, Ti^{4+} and Zr^{4+} [41, 42, 43, 44]. The immobilized metal affinity chromatography (IMAC) [45] as well as the metal oxide affinity chromatography (MOAC) [46] have become the two most popular methods for the enrichment of phosphorylated peptides [40]. In IMAC, metal ions are immobilized on resins coated with iminodiacetic acid (IDA) or nitrilotriacetic acid (NTA). Despite its wide use, the method comes with certain drawbacks; particularly, the low tolerance towards buffers and salts and unspecific binding of acidic peptides [47].

The enrichment of phosphopeptides with MOAC is based on the complex formation between phosphorylated peptides and metal oxides and hydroxides. The most widely used in MOAC is TiO_2 due to its superior sensitivity and specificity over other metal oxides [48]. Like IMAC, MOAC also suffers from the co-enrichment of acidic peptides. Specificity can be improved by using 2,5-dihydroxybenzoic acid (DHB) solution containing acetonitrile and trifluoroacetic acid for loading of samples [49]. Elution from the columns is performed under basic conditions of approximately pH 10.5 using ammonium hydroxide.

A difference that has been observed between these two methods is that IMAC favors multiply phosphorylated peptides while MOAC yields preferably single-phosphorylated peptides [47]. This fact has led to the development of workflows which aim to complement these characteristics [50].

1.3.2 Enrichment of phosphotyrosine peptides

Afore mentioned methods using chemical resins enrich the global phosphoproteome which is, as mentioned in Section 1.1, largely dominated by phosphorylated threonine and, especially, serine residues. Investigation of tyrosine phosphorylation, which plays essential roles in proliferation, metabolism, differentiation and survival, remains a struggle due to its highly dynamic and short-lived nature [51]. The specific and sensitive enrichment of tyrosine-phosphorylated peptides requires an own set of methods. For this enrichment, immunoaffinity methods like immunoprecipitation are usu-

ally employed [40, 52, 53], using antibodies developed to bind phosphorylated tyrosine residues [54, 55]. In immunoprecipitation, these antibodies are immobilized on protein A- or G-coated magnetic or agarose beads. Peptides bound to the antibodies are pulled down by centrifugal force or on a magnetic rack and subsequently eluted from the beads.

As an alternative to antibody-based approaches, a recent study used a SH2-domain derived, so-called superbinder as an affinity agent to enrich phosphotyrosine peptides, yielding a greater depth in the coverage of the phosphotyrosine proteome than the antibody-based enrichment [56].

Deep phosphoproteomic analysis often combines the enrichment of the global proteome and tyrosine phosphorylated peptides to gain better coverage of the phosphoproteome [57, 58]. Other studies focus specifically on tyrosine phosphorylation and facilitate important insight into signaling networks that are largely driven by the dynamics of tyrosine phosphorylation [59, 60, 53].

1.4 Mass Spectrometry for Proteomics

Mass spectrometry (MS) is an instrumental method of analytical chemistry that measures the mass-to-charge (m/z) ratio of gas-phase ions. In a mass spectrometer, an analyte molecule is ionized, separated in an electro-magnetic field and subsequently detected based on the trajectory or time of flight of the ion [61].

MS-based proteomics is state-of-the-art in deciphering the inventory of a cell's proteins. The three main applications for MS in proteomics are studies of protein expression, protein interaction and sites of protein modification [62]. A typical workflow starts with the extraction of whole cell protein by lysis. Subsequent use of the endopeptidase trypsin cleaves specifically C-terminal of lysine and arginine residues, digesting the proteins into peptides. Finally, the chromatographic separation of peptides and subsequent mass spectrometry analysis reveals the protein complement of a cell's genome [63].

1.4.1 Soft ionization techniques

While previous ionization technologies resulted in the destruction of biological compounds like proteins and nucleotides, the soft ionization techniques — named after their characteristic to cause very little fragmentation of the analyte — matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) paved the way of MS for the studies of proteins [64]. In ESI, the analyte is dissolved in a solvent and sprayed by a thin needle into the spectrometer. A high electrical potential is constantly applied to the needle, resulting in the nebulization of charged droplets. Depending on the polarity of the applied voltage, the droplets are either positively or negatively charged. Subsequently, the droplets are reduced in size by evaporation, until a critical point is reached at which the so-called coulomb explosion ejects charged ions into the gaseous phase [65]. MALDI starts with the co-crystallization of the analyte-solvent into a solid phase matrix. The sample is therefore mixed with excessive amounts of a solution of a small organic compound and then air-dried [66]. Application of a short laser pulse of high intensity leads to the desorption and ionization of the analytes in the matrix [64].

Both ESI and MALDI can be regarded as highly sensitive techniques for ionization.

The main difference is the state at which the sample is introduced into the instrument, being solvated in ESI or in a solid state in MALDI. ESI can be easily coupled to liquid chromatography and enables an efficient workflow for quantitative measurements [67]. In the course of a quantitative ESI experiment it is necessary to incorporate an internal standard to compensate for losses in the sample preparation process and the sensitivity of the intrument [65]. MALDI has also been used for quantitative measurements [68] and was coupled to liquid chromatography [69]. However, MALDI does not yield a uniformly distributed representation of the analyte and the quality of MALDI spectra can be significantly influenced by the position of the laser beam. This irreproducibility poses difficulty for such applications [64, 70].

1.4.2 Fragmentation

After ionization of the analyte, it is necessary to create fragments of the peptides for their characterisation via peptide sequencing. The first fragmentation method to be developed was collision-induced dissociation (CID) in 1977 [71]. Fragmentation of the peptides is achieved as a result of collision of the analytes with an inert neutral gas, such as helium, nitrogen or argon. The kinetic energy of the collision leads to disruption of bonds along the peptide-backbone [72]. The breakage of the amide bond results in so-called b- and y-ions, describing the resulting fragments beginning from the N- and C-terminus respectively [73]. The fragmentation pattern of this method may depend on several factors like amino acid composition, size and charge of the peptide [72]. Moreover, it has been repeatedly shown that fragmentation generally does not occur randomly but follows certain patterns that have been revealed by data-mining and statistical analysis [74, 75, 76]. A typical occurrence during fragmentation with CID in phosphopeptides is the neutral loss of phosphate due to the labile phosphoester bond relative to other bonds in the peptide [77]. This may lead to CID spectra of phosphopeptides that are largely dominated by a single peak corresponding to the neutral loss of H_3PO_4, hindering the correct identification of peptide fragments and the localization of the phosphorylation site [78]. This affects phosphotyrosine residues to a lesser extent, as the C-O bond is stabilized by the aromatic ring of the tyrosine [79].

In order to obtain more sequence information in CID spectra with neutral loss, the fragments can undergo a second stage of activation that reveals the fragments of the

peptide backbone and thereby the amino acid sequence. This so-called MS^3 method is usually employed in a data-dependent setting on spectra with an intense peak indicating neutral loss [72]. A downside of this approach is the considerable increase of analysis time per peptide. Multistage Activation (MSA) can be applied to reduce this analysis time while still achieving a similar result. For this approach the second isolation step is omitted and the neutral loss ion is activated with the precursor ion fragments still present. This results in an information-rich, composite spectrum of MS/MS and MS^3 fragments [80]. MSA is most suited in samples that are rich in phosphopeptides, which is the case in a typical phosphoproteomic experiment after enrichment of phosphopeptides.

To improve upon the sensitivity of traditional, regular CID, higher-energy C-trap dissociation (HCD) has been developed for the linear trap quadrupole orbitrap (LTQ Orbitrap) MS platform. This method provides better resolution, accuracy and less loss, than CID, especially with lower masses [81]. It is debated whether CID-based MSA or HCD are more suitable for application in phosphoproteomics and studies show differing findings [82, 83, 84].

Another approach in terms of fragmentation methods are electron-based dissociation methods that can leave PTMs largely intact during fragmentation [85]. Electron capture dissociation (ECD) uses low-energy electrons that interact with the protonated precursor ion to form a radical cation, leaving the peptide in a transient state that ultimately produces fragments [86]. Linear ion trap CID is generally more efficient and yields more phosphopeptide identifications, whereas ECD is more useful for the localization of phosphorylation in peptides. In a study conducted by Sweet et al., ECD showed an overall poorer performance, even though it used a Fourier-transform ion cyclotron resonance (FT-ICR) MS, a high resolution instrument compared to the linear ion trap MS used with CID [87].

Another factor which renders ECD less practicable than other fragmentation methods is the high cost of FT-ICR instruments. A similar adaptation of ECD for less costly forms of MS is the electron-transfer dissociation (ETD) that was developed a few years later [88]. ETD produces radical anions from molecules with low electron affinity that pass on the electron to the precursor peptide. This results in the formation of a radical cation and subsequent fragmentation, similar to ECD.

1.4.3 Mass analyzers

Several types of mass analyzers are typically used in proteomics workflows. In linear ion traps (LIT), ions are radially and axially confined by a two-dimensional radio-frequency field as well as static potentials applied to electrodes at the end of the quadrupoles. Among their advantages are high storage capacity, the rather simple construction as well as their high injection efficiency [89]. Orbitraps are another example of trapping type mass analyzers that confine ions by balancing the electrostatic attraction to the inner electrode with the inertia of the ion, resulting in a characteristic trajectory around the central electrode and oscillations that are proportional to the m/z ratio [90]. This analyzer gives a very high resolution and high mass accuracy while still being comparably small and cheap.

Another type of mass analyzer that uses m/z dependent trajectories for selection is the quadrupole mass analyzer in which ions are separated by oscillating electric fields applied to the two pairs of rods. The major advantage of this analyzer is its low cost, small size and maintainability compared to other methods [64].

Time of Flight (TOF) analyzers utilize the proportionality of the flight time of ionized molecules to the detection unit, upon acceleration in an electric field. In order to correct for variance in the distribution of kinetic energy of the ions and to thereby increase the resolution of the instrument, reflectron TOF analyzers are usually employed which use an electrostatic ion mirror to further extend the travelling path of the molecules [64].

Today, MS for proteomics is almost always used in a hybrid setting, coupling multiple mass analyzers for gains in specificity and sensitivity as well as for being able to process the large amount of peptides that are being introduced into the MS in short time. Many so-called tandem MS (MS/MS) instruments combine, for example, quadrupole and TOF (qTOF), quadrupole and orbitrap or LIT and orbitrap systems [91, 92, 93].

1.4.4 Quantitative Proteomics

For many types of proteomics studies, the quantification of proteins is the central aspect of the underlying question. However, this task is also one of the technically most challenging. Several strategies can be employed for quantitative proteomics: (a) *in*

vivo metabolic labelling with stable isotopes, (b) chemical labelling of peptides with chemical mass or isobaric tags or (c) label-free approaches.

In Stable Isotope Labelling by Amino acids in Cell culture (SILAC), cultivation media is usually supplemented with ^{13}C or ^{15}N isotopes, which leads to the incorporation of these isotopes into the metabolism of the cells. This change of mass of the peptides can be detected by the MS and can be used for quantitation. Metabolic labelling is among the quantitative methods with the highest accuracy since samples can be mixed at the stage of living cells already [94]. However, this labelling strategy is limited to comparisons of just 2–3 states or samples.

With Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) and Tandem Mass Tags (TMT), two methods for labelling peptides with isobaric tags are available that give rise to characteristic reporter ions. This ways, up to sixteen different samples in TMT [95, 96] or eight in iTRAQ can be quantified in a multiplexed analysis [97]. The major limitation of isobaric labels arises with the co-fragmentation of peptides in the same mass window as the precursor of interest [98].

In the label-free scenario, relative quantitative comparisons are achieved via calculating the area under the curve (AUC) of peaks in the spectrum or by spectral counting. The lack of labelling agents aids in flexibility in terms of experimental design, avoids additional workflow steps and reduces implementation cost [99]. The samples in label-free analyses are not mixed, which means that the number of samples or conditions is not restricted, but also that the accuracy of the quantification is lower in the overall experimental process, compared to labelled approaches [94]. Furthermore, an elongated data acquisition time has to be considered in label-free experiments. The majority of challenges arise in the post-acquisition analysis and robust quantitation requires specialised proteomics software that is improved continuously [100, 101, 102].

1.4.5 Bioinformatics

Post-acquisition analyses are done computationally and are largely carried out in automated and standardized pipelines [103]. The first steps in the analysis of the raw data includes the identification of peptides. The SEQUEST algorithm was the first to be developed and identifies peptides spectra by correlating them to *in silico*-generated

peptide spectra from a protein database [104]. Similarly, MASCOT software uses a probabilistic approach to compute a score that indicates the match between the experimental spectrum and the theoretical spectra [105].

These database searching algorithms report several scores that are used to describe the confidence of each peptide match. Reports have shown a significant overlap between between true positive and false positive peptide matches in these scoring methods [106, 107]. In order to minimize false discoveries, the use of decoy protein databases and score thresholds improves the rate of true positive identifications. Instead of manually setting these thresholds and creating decoy databases for each specific application, the Percolator algorithm uses a semi-supervised machine learning approach to discriminate between correct and incorrect peptide spectrum matches (PSMs) [108].

The aforementioned programs are not optimized for identification of PTMs. Specialized software such as phosphoRS is used to calculate individual site probabilities for phosphorylation in peptides [109]. The algorithm scans each MS/MS spectrum in windows of 100 m/z and calculates the cumulative binomial probability for each theoretical isoform representing phosphorylations. Notably, this software can be used with spectra from any fragmentation method including MSA, HCD and ETD.

2 Project aim

The aim of this study is

- to develop a protocol for the specific enrichment and analysis of peptides containing phosphorylated tyrosine residues from biopharmaceutical production cell lines such as Chinese Hamster Ovary cells,
- to elucidate the role of different common methods in phosphoproteomics in terms of quantity and quality of discovered tyrosine-phosphorylated peptides and
- to characterize proteins that could be retrieved in these analyses, as well as,
- to describe the molecular functions, cellular localization and biological processes that are enriched by this workflow using Gene Ontology analyses.

3 Materials and Methods

3.1 Cell Culture

CHO K1 cells were grown in 250 ml sterile disposable flasks in a culture volume of 20 ml with CD CHO medium (Gibco, Cat.-Nr. 10743029) supplied with 2 µl ml^{-1} anti-clumping agent, incubated at 37 °C, 5 % CO_2 and 80 % humidity at 170 rpm in a shaking incubator (ISF1-XC Climo-Shaker, Kuhner). Culture density was monitored by cytometry using a Guava EasyCyte HT (Merck, GER). Cells were split to 0.2×10^6 cells ml^{-1} every 3–4 days.

3.2 Cell Harvesting

The culture volume was transferred to a universal conical tube and centrifuged at 1000 rpm for 5 min to pellet the cells. The cell pellet was washed once with 5 ml ice-cold PBS and subsequently transferred in to a 1.5 ml microcentrifuge tube 1 ml ice-cold PBS. Afterwards, PBS was removed by centrifugation at 5000 rpm and 4 °C for 5 minutes and the remaining cell pellet was stored at −20 °C.

3.3 Cell Lysis

The pellet was thawed and resuspended in 200 µl lysis buffer (8 M Urea, 50 mM Tris, 75 mM NaCl, pH 8.2). Subsequently, 100 × HALT Phosphatase Inhibitor Solution (Cat.-Nr. 78420, Thermo Fischer) was added to the suspension. The lysis was conducted by ultrasonication with three pulses, each with the duration of 30 seconds and with 5 minutes pause in between. Afterwards, the lysed sample was centrifuged at 14 000 rpm and 4 °C for 15 minutes. The supernatant was transferred to a new 1.5 ml microcentrifuge tube and total protein was quantified.

3.4 Protein Quantitation

Protein quantitation was carried out by a Bradford assay (BioRad). Bovine serum albumin (BSA) standard dilution series was made from a stock solution of $2\,\mathrm{mg\,ml^{-1}}$ in a 2-fold manner. The sample was diluted in order to fit approximately into the standard's range of $0.25\,\mathrm{mg\,ml^{-1}}$ to $1\,\mathrm{mg\,ml^{-1}}$. 5 µl of each sample and standard were transferred into wells of a 96-well plate. Subsequently, 250 µl of Bradford dye were added. Measurements were conducted in triplicates, using a Multiskan GO instrument with corresponding SkanIt software (v3.2.1.4, Thermo Fischer, USA) at 595 nm after an incubation of 10 minutes at room temperature.

3.5 In-solution Digestion

The sample was split into aliquots of 1 mg of protein and dithiothreitol (DTT) was added to a final concentration of $5\,\mathrm{mmol\,l^{-1}}$ and incubated at 56 °C for 25 minutes. Afterwards, samples were cooled to room temperature. $0.5\,\mathrm{mol\,l^{-1}}$ freshly prepared iodoacetamide was added to the mix to a final concentration of $14\,\mathrm{mmol\,l^{-1}}$ and incubated for 30 minutes at room temperature in the dark. Subsequently, additional DTT was added to make a final concentration of $10\,\mathrm{mmol\,l^{-1}}$ in the mix. Thereafter, it was incubated for 15 minutes at room temperature in the dark. The sample was transferred to a 2 ml microcentrifuge tube and diluted 1:5 with $25\,\mathrm{mmol\,l^{-1}}$ Tris-HCl, pH 8.2. Tryptic digest was conducted by adding 1 µg trypsin per 50 µg substrate and incubating at 37 °C and 200 rpm for 4 hours using sequencing grade trypsin (Promega). Afterwards, an additional 1 µg trypsin per 100 µg protein was added to the mix and incubated at 37 °C with shaking at 200 rpm overnight.

Digested samples were cooled to room temperature and LC-MS grade trifluoracetic acid (TFA) was added to a concentration of 0.4 % (v/v) and checked for a $\mathrm{pH} < 2$. Subsequently, samples were centrifuged at $2500 \times \mathrm{g}$ at room temperature for 10 minutes. The supernatant was transferred to a fresh 2 ml microcentrifuge tube, the pellet was discarded. Samples were stored at −20 °C.

3.6 Peptide desalting

First, Sep-Pak C18 cartridges (Cat.-Nr. WAT054945, Waters) were washed and conditioned with 6 ml acetonitrile, followed by 3 ml 50 % acetonitrile, 0.5 % acetic acid. Subsequently, the column was equilibrated with 9 ml 0.1 % TFA. The sample was loaded onto the column and reapplied two times. Washing and desalting was conducted with 6 ml 0.1 % TFA. Finally, TFA was removed by washing with 900 µl acetic acid. Desalted peptides were eluted with 5 ml of 50 % acetonitrile, 0.5 % acetic acid into a 15 ml conical tube and snap-frozen with liquid nitrogen. The sample was lyophilized overnight and afterwards stored at −20 °C.

3.7 Phosphopeptide-Enrichment with TiO_2

Enrichment of phosphopeptides was conducted using High-Select™ TiO_2 Phosphopeptide Enrichment Kit (Cat.-Nr. A32993, Thermo Fischer). The TiO_2 spin tip and adaptor was placed in a 2 ml microcentrifuge tube. 20 µl of wash buffer was applied to the column followed by a centrifugation at 3000 × g for 2 minutes before equilibrating using binding buffer and centrifuging again at 3000 × g for 2 minutes. The filtrates were discarded. Samples were resuspended in 150 µl of binding buffer and mixed.

The equilibrated TiO_2 tip was transferred to a fresh 2 ml microcentrifuge tube and the resuspended sample was applied to the column. After centrifugation at 1000 × g for 5 minutes, the flow-through was reapplied to the column and spun again. After putting the spin column into a new 2 ml microcentrifuge tube, the bound sample was washed by adding 20 µl of binding buffer and centrifugation at 3000 × g for 2 minutes. A final wash was conducted with 20 µl of wash buffer and centrifugation as before. Washing was repeated once in the same order of buffers.

Finally, the column was washed with 20 µl of LC-MS grade water and centrifuged at 3000 × g for 2 minutes. After placing the TiO_2 spin column into a new 1.5 ml microcentrifuge tube, the sample was eluted from the column by the addition of 50 µl elution buffer and centrifugation at 1000 × g for 5 minutes. This elution step was repeated once and eluates were pooled.

In order to remove the acidic solution from the eluate, samples were concentrated

in vacuo and used for subsequent immunoprecipitation for analysis with, or C18 purification for analysis without phosphotyrosine enrichment.

3.8 Immunoprecipitation

Per precipitation, 100 µl, corresponding to 3 mg of magnetic Dynabeads were resuspended by vortexing and transferred into a new 1.5 ml microcentrifuge tube. Antibody solutions were prepared by mixing 8 µg of the respective antibody — P-Tyr-1000 (Cat.-Nr. 8954, Cell Signaling) or 4G10 (Cat.-Nr. 05-321, Sigma Aldrich) — per 1 mg of peptide (before phosphopeptide enrichment) in 200 µl antibody binding & washing buffer.

The supernatant of bead-containing tubes was discarded by the use of a magnetic rack, before resuspending the beads in the antibody solution. For conjugation of the antibodies to the magnetic beads, this mixture was incubated rotating at room temperature for 1 hour.

Antibody-conjugated beads were washed once with 200 µl of antibody-binding & washing buffer by gentle up-and-down pipetting and removal of the supernatant on the magnetic rack. Finally, the beads were resuspended with 1 mg of peptide and incubated rotating at 4 °C overnight.

After antibody binding, the unbound supernatant was transferred to a 1.5 ml microcentrifuge tube and stored at −20 °C for further experimentation. The phosphotyrosine-bound beads were washed three times with 200 µl wash buffer. During a final wash with 100 µl wash buffer, the beads were transferred to a new 1.5 ml microcentrifuge tube and the supernatant was removed on the magnetic rack. Subsequently, phosphotyrosine peptides were eluted from the beads with either 50 µl 60 % acetonitrile/0.1 % TFA or with 50 µl of 100 mM glycine at pH 2, by incubation rotating at room temperature for 3 minutes. The elution step was repeated once and supernatants were pooled in a new 1.5 ml microcentrifuge tube. Phosphotyrosine peptide-containing samples were concentrated *in vacuo* and afterwards stored at −20 °C.

3.9 C18 Purification

Samples from immunoprecipitation or TiO_2 phosphopeptide enrichemnt were resuspended in 100 µl of 2 % acetonitrile/0.1 % TFA.

The C18 resin of the spin column was activated using 200 µl 50 % acetonitrile and centrfigation at 1500 × g for 1 minute. The supernatant was discarded and the resin was equilibrated twice with 200 µl 2 % acetonitrile/0.1 % TFA.

Afterwards, the sample was loaded on top of the resin and the column was centrifuged at 1500 × g for 1 minute. The flow-through was reapplied once to ensure complete binding of the sample to the resin. Washing was conducted twice with 200 µl 2 % acetonitrile/0.1 % TFA and centrifugation at 1500 × g for 1 minute.

The column was placed in a new 1.5 ml receiver tube and the peptides were eluted with 20 µl 70 % acetonitrile. After a short incubation time of 1–2 minutes, the column was centrifuged at 1500 × g for 1 minute. The elution step was repeated once and filtrates were pooled into the same receiver tube. The solute was removed thereafter *in vacuo*. Subsequently, the pellet containing the purified peptides was stored at −20 °C or used immediately for mass spectrometry.

3.10 Preparation for Mass Spectrometry

Peptides were resuspended in 22 µl 2 % acetonitrile/0.1 % formic acid and vortexed for several seconds. The mixture was incubated at room temperature for 20 minutes. Afterwards, the sample was sonicated in the sonication bath for 5 minutes. The amount of peptides was measured in the NanoDrop (Thermo Fischer). 20 µl were transferred into a vial and put in the sonication bath for 5 minutes.

3.11 LC-MS/MS

The analysis was performed using a UltiMate 3000 nanoRSLC coupled in-line with Orbitrap Fusion Tribrid mass spectrometer (both Thermo Fischer, USA). 1 µg of peptide per samples were loaded onto the trapping column C18 PepMap, 300 µm × 5 mm

(Thermo Fischer, USA) and desalted for 3 minutes at a flow rate of 25 µl min^{-1} using 0.1 % (v/v) TFA/2 % (v/v) acetonitrile. Afterwards, peptides were resolved on an Easy-Spray C18 75 µm × 250 mm column with a bead diameter of 2 µm (Thermo Fischer, USA). Solution A for the gradient consisted of 0.1 % (v/v) formic acid. Solution B consisted of 80 % (v/v) acetonitrile and 0.08 % (v/v) formic acid. The gradient, starting from 98 % of solution A and 2 % solution B, was raised to 35 % solution B over the course of 120 minutes with a flow rate of 300 nl min^{-1}.

MS1 spectra were acquired over 400–1400 m/z in the Orbitrap at 120 K resolution and 200 m/z. Automatic gain control was set to 4×10^5 ions with a maximal injection time of 50 ms. Tandem MS analysis was conducted in data-dependent mode with normalized collision energy optimized at 35 % for Collision Induced Dissociation. Spectra for MS2 were acquired in the ion trap. The intensity threshold for fragmentation was set to 10 000 and included the charge states +2 to +6. The dynamic exclusion was set to 50 s and mass tolerance to 10 ppm. Fragmentation using Multistage Activation (MSA) was triggered for any precursor ion with neutral-loss of mass 97.9763 with a tolerance of 0.5 m/z. The automatic gain control was set to 20 000 with a maximal injection time of 90 ms.

3.12 Computational Analysis of LC-MS/MS data

Analyses were conducted using the Proteome Discoverer software (v2.1, Thermo Fischer, USA), which incorporates the algorithms used for analysis. Identification of proteins was conducted using the SEQUEST HT algorithm [104] with the NCBI Chinese Hamster Ovary (*Cricetulus griseus*) protein database from November 2017, containing 24 906 sequences. The fragment mass tolerance was set to 0.6 Da and precursor mass tolerance to 10 ppm. Enzyme specificity was set to Trypsin with two missed cleavages allowed. Carbamidomethylation of cysteine was set as static modification. Phosphorylation of serine, threonine and tyrosine and oxidation of methionine were set as dynamic modifications. Identification of phosphorylated sites was conducted using the phosphoRS algorithm [109].

Data were filtered to a false discovery rate (FDR) of 1 % by automatic decoy searching in SEQUEST and by phosphoRS probability scores for phosphorylated sites by > 75 %.

3.13 Gene Ontology

Tyrosine-phosphorylated proteins were assigned to equivalent mouse gene symbol identifiers by using the most similar mouse protein, with an identity of at least 90 %, from the UniProtKB database. Gene Ontology term enrichment was conducted using the DAVID web interface [110]. Terms for cellular components, biological processes and molecular functions, as well as KEGG pathway analysis were subject to an adjusted p-value (Benjamini-Hochberg, [111]) cutoff of < 0.05.

4 Results

4.1 Optimization of Phosphotyrosine Enrichment

The extreme sub-stoichiometric nature of tyrosine phosphorylation triggers it difficult to achieve an appropriate coverage of the tyrosine phosphoproteome. Specific enrichment of tyrosine-phosphorylated peptides by immunoprecipitation with anti-phosphotyrosine antibodies is the most widely applied method to improve the number and confidence of hits.

4.1.1 Comparison of antibodies

Figure 2 shows initial experiments that were conducted in order to improve the number of detected phosphotyrosine sites. Two antibodies, 4G10 and P-Tyr-1000, were available which are both regularly applied for the enrichment of phosphotyrosine in proteomics experiments. A study with colorectal cancer cells, conducted by van der Mijn et al. [112], found that P-Tyr-0000 performed superior to 4G10. However, this study used a very large quantity of cells and no concentration of P-Tyr-1000 is reported. To evaluate the performance of both antibodies on a smaller amount of CHO cells, this study was tried to be replicated using the antibody P-Tyr-1000 according to manufacturer's recommendations for immunoprecipitation. The result of this initial experiment can be seen as experiments A, for P-Tyr-1000, and B, for 4G10, in Figure 2.

Using the antibody 4G10 and MSA fragmentation gave rise to only two detected phosphotyrosine residues from 1 mg of whole cell lysate. The antibody P-Tyr-1000 is not provided with information regarding its concentration. Hence it was used with a dilution of 1:200 according to manufacturer's recommendations. This did not result in the detection of any phosphotyrosine sites which is likely due to a too low amount of antibody used for immunoprecipitation. Correspondence with the manufacturer after the experiment revealed the concentration of the antibody to be $928\,\mu g\,ml^{-1}$. In follow-up experiments, the antibody P-Tyr-1000 was used as described in Section 3.8. The number of identified phosphotyrosine sites increased in subsequent trials. Experiments C, D and E were conducted with preceding enrichment of phosphorylated peptides using

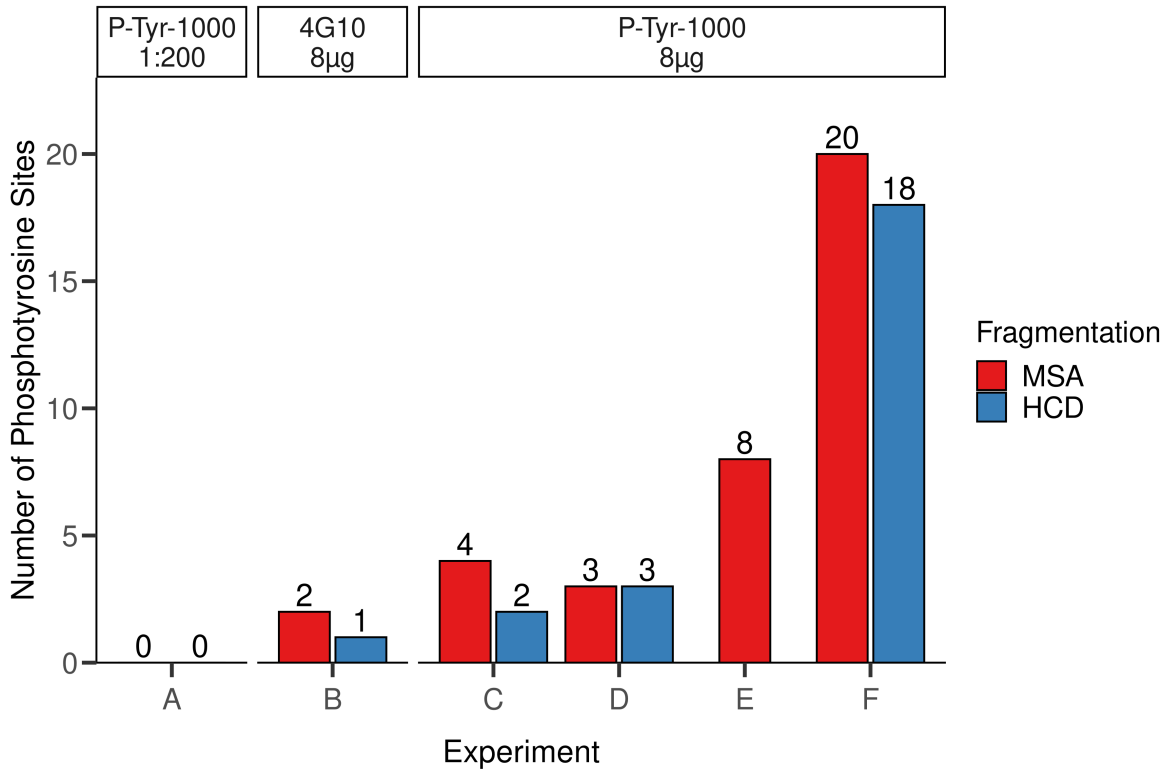

Figure 2: Improvements in the number identified phosphotyrosine sites throughout initial experiments. Experiments are denoted from A–F in chronological order. Colored in red and blue are fragmentations methods MSA and HCD, respectively.

TiO_2 enrichment. For experiment E, only MSA fragmentation has been used. Experiment F was conducted without prior enrichment of phosphopeptides but by applying immunoprecipitation directly with whole cell peptide.

Literature research revealed different approaches to immunoaffinity enrichment of phosphotyrosine. Some studies enriched global phosphopeptides using MOAC or IMAC before enriching specifically for phosphotyrosine [113, 14, 114], while others enriched from whole peptide instead [115]. Both approaches were tested in initial experiments. In experiment F, conducted without prior enrichment via TiO_2, more phosphotyrosine sites could be discovered compared to other experiments.

Fragmentation with MSA enabled the discovery of more phosphotyrosine sites in most experiments. However, a clear comparison cannot be carried out from these experiments due to the low number of identified sites.

Implications of fragmentation methods as well as preceding enrichment of global

phosphopeptides on the quantity and quality of identified phosphotyrosine sites were further elucidated in a subsequent experiment reported in Section 4.3.

4.1.2 Antibody depletion

A major challenge in the analysis of phosphorylation events by proteomics is the inevitable peptide loss during the handling and processing of the sample. This is especially critical when the aim is to enrich phosphotyrosine — the least abundant type of phosphorylation. An important step in an immunoaffinity enrichment procedure is the depletion of the antibody from the sample. Abe et al. have conducted a study in which they tried to optimize several parameters of the phosphotyrosine enrichment procedure [113]. They showed that by using Fe^{3+}-IMAC for depletion of antibodies after immunoprecipitation, considerably more phosphotyrosine sites can be discovered than by using C18 columns. To elucidate whether the same can hold true for MOAC with TiO_2, measurements of the peptide concentration in the samples are used as a proxy for recovery of peptide from antibody depletion methods. Measurements were conducted after immunoprecipitation as well as after conducting either C18 purification or TiO_2 MOAC for depletion of antibodies using a NanoDrop One (Thermo Fischer, USA) for the peptide quantification.

Table 1: Loss of peptide caused by the respective processing steps based on peptide measurements before and after each step.

Method	Mean % Loss	SD
C18 Purification	33.26	1.78
TiO_2 enrichment	63.55	6.24

Table 1 shows the percent loss of peptide from the depletion of the antibody by the respective methods. C18 cartridges recover significantly more peptide than using TiO_2. Therefore, the C18 was used for the remainder of the project.

4.2 Antibody binding bias

Anti-phosphotyrosine antibodies have previously been reported as sensitive towards sequence context [116]. Amino acids in the vicinity of a phosphorylated tyrosine residue in the peptide may encourage or prevent binding of anti-phosphotyrosine antibodies. Studies of the phosphotyrosine proteome of mammalian cell factories require a reliable and sensitive method for enrichment of phosphotyrosine peptides. Therefore, the frequencies of amino acids around the binding site of the antibody P-Tyr-1000 in 43 peptide sequences were analyzed and are illustrated in a weblogo.

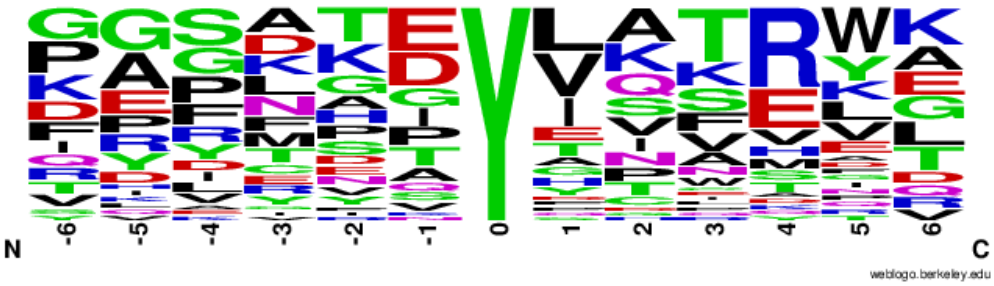

Figure 3: Frequency of amino acids in the vicinity of bound phosphotyrosine residues. Visualized using WebLogo [117].

Figure 3 shows the frequency of amino acids around the identified tyrosine phosphorylation sites. One notable observation in the frequency of amino acids is the occurrence of hydrophobic amino acids leucine, valine and isoleucine at +1. Reports show a very strong enrichment of valine at position +1 of tyrosine phosphorylation sites [118]. Other hydrophobic amino acids, however, are not strongly enriched at this position and their frequency may be attributed to a bias in antibody binding. Tinti et al. have shown in peptide microarrays that hydrophobic amino acids, leucine in particular, are enriched in the vicinity of the antibody binding site of phosphotyrosine containing peptides [116].

The position −1 shows higher frequencies of acidic amino acids glutamic acid and aspartic acid while Tinti et al. report depletion of acidic amino acids in the sequence context of anti-phosphotyrosine antibodies. Arginine, here at position +4, is often found enriched in sequences downstream of the antibody binding site.

In general, however, it can be seen from these experiments that the anti-phosphotyrosine antibody P-Tyr-1000 does not show any strong bias with regards to sequence context.

4.3 Comparison of protocols

Literature research revealed a wide variety of methods and protocols for the enrichment of phosphotyrosine peptides with immunoprecipitation. Most previous studies are not using label-free mass spectrometry and many use epidermal growth factor (EGF) treatment for stimulation of phosphorylation [113, 119, 115]; both of which are not relevant for the tasks underlying this thesis.

Here, different methods for enrichment of phosphotyrosine peptides with immunoprecipitation are compared. First, it was investigated whether it is advantageous to first enrich for all phosphopeptides by TiO_2-enrichment, with the rationale that enrichment of phosphorylated tyrosine residues by antibodies could be more effective after being concentrated, due to the substochiometric nature of tyrosine phosphorylation. On the other hand, however, each additional processing step also causes more loss of peptides of interest.

Advances in mass spectrometry lead to enhancements in resolution and depth of proteomic analyses. Enrichment methods were developed to cope with the lesser capabilities of older instruments while the benefit of some enrichment methods with current devices could be debated. Hence, in addition to aforementioned enrichment strategies, the specific enrichment of phosphotyrosine peptides is omitted in another protocol to investigate the benefit of the immunoprecipitation.

In previous studies, different elution methods were used in immunoprecipitation: Elution by acetonitrile and trifluoracetic acid (e.g. in [113]) as well as elution by low-pH glycine (as in [119]). In order to compare the performance of these two elution methods, both were applied and analysed.

Furthermore, both MSA as well as HCD are regularly applied for proteomics and phosphoproteomics. Samples were analyzed with both fragmentation methods to gain another layer of insight into implications of discovered tyrosine phosphorylations.

An overview of the experimental setup used to elucidate these aspects can be seen in Figure 4. After the digestion of the proteins into peptides and the subsequent desalting, as explained in Sections 3.5 and 3.6, the protocols diverge into three variations that aim to:

- enrich peptides with phosphorylated tyrosine residues using immunoprecipitation from whole peptide samples,
- enrich peptides with phosphorylated tyrosine residues using immunoprecipitation from afore, by TiO_2, enriched phosphorylated peptides, or
- enrich phosphorylated peptides by TiO_2, equivalently to a regular phosphoproteomics workflow.

The experiment was carried out with 5 mg of raw peptide and split evenly across all samples, amounting to 1 mg per sample, after the desalting step to ensure equal starting material in each protocol. Immunoprecipitations in these experiments are conducted with elution by acetonitrile (ACN)/TFA or low-pH glycine (GLY) as explained in Section 3.8. Each sample derived from this experimental workflow was analyzed by mass spectrometry using the fragmentation methods MSA as well as HCD.

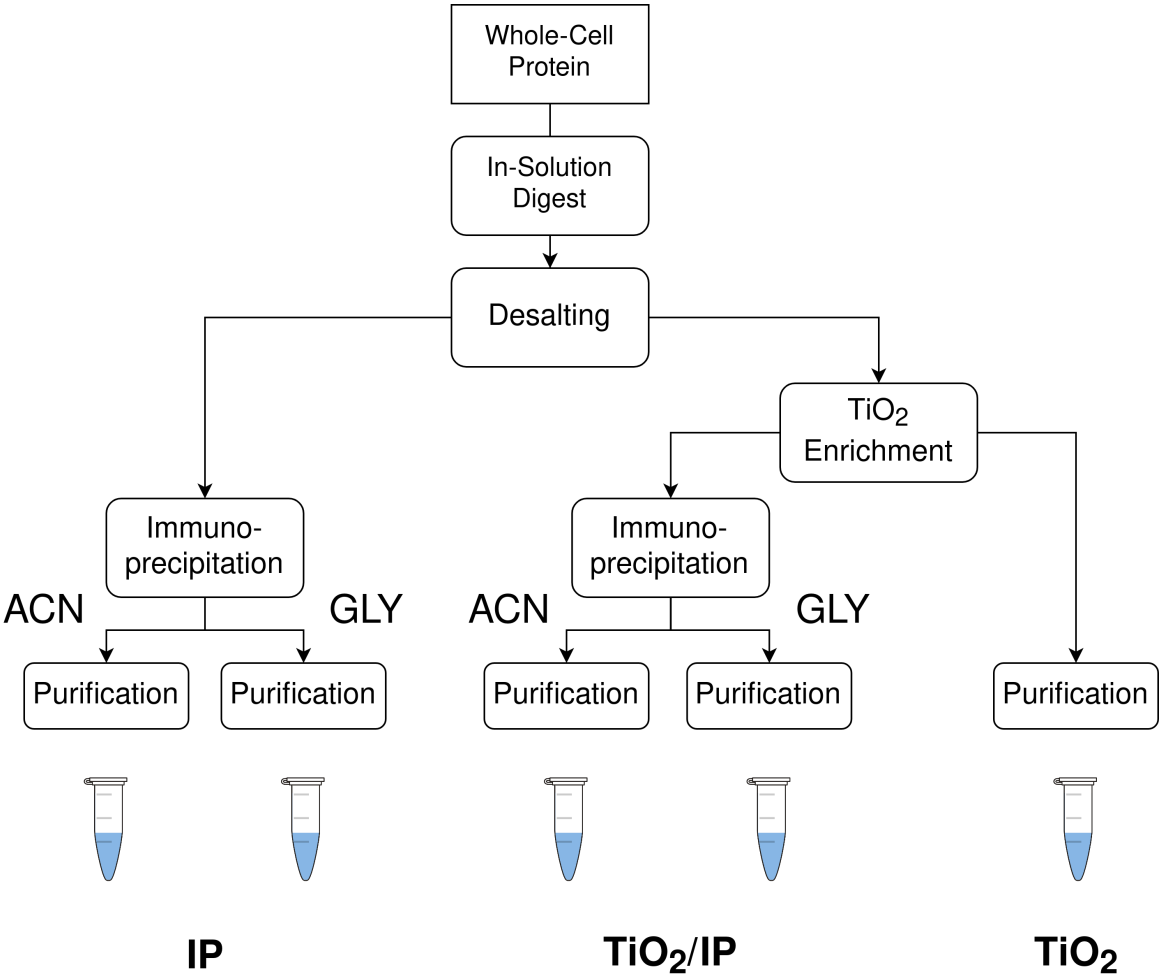

Figure 4: Experimental setup for the comparison of protocols for quantitative and qualitative analysis of phosphorylated tyrosine residues in proteomes.

4.4 Identified phosphotyrosine

Table 2 shows the amount of peptides and phosphorylated sites that were derived from each protocol variation. The regular phosphoproteomics workflow that omitted any specific enrichment of phosphorylated tyrosine yields a typical number and distribution of discovered sites that reflects the low abundance of tyrosine phosphorylation.

Protocols involving immunoprecipitation lead to very few discovered overall peptide and phosphorylated sites compared to the usual workflow. Protocols that used low-pH glycine did not yield any phosphotyrosine sites and only very few peptides in general, clearly showing the inferior performance of this elution method in comparison to ACN/TFA. Furthermore, a greater depth is gained by fragmentation of the sample by MSA, as opposed to HCD, which is reflected by greater numbers of peptide and phosphorylated residues discovered in every protocol. The increase is drastic in the case of the IP ACN-eluted protocol, with approximately ten times more peptide and phosphorylated sites discovered by fragmentation with MSA. The regular phosphoproteomics workflow benefitted from this fragmentation as well, by doubling in numbers.

Table 2: Number of discovered total and phosphorylated peptides as well as number of phosphorylated sites with phosphoRS probabilities > 75 %. Phosphorylated sites are further subdivided into serine, threonine and tyrosine sites

	MSA				
	ACN		GLY		
	IP	TiO_2/IP	IP	TiO_2/IP	TiO_2
Total Peptide	306	45	69	129	6496
Phos. Peptide	37	11	1	3	5093
Phos. Sites	39	11	1	4	6480
pS	5	0	1	3	5948
pT	2	0	0	1	496
pY	32	11	0	0	36
	HCD				
	ACN		GLY		
	IP	TiO_2/IP	IP	TiO_2/IP	TiO_2
Total Peptide	38	15	12	7	3646
Phos. Peptide	3	2	0	0	2776
Phos. Sites	3	2	0	0	3371
pS	0	0	0	0	3165
pT	0	0	0	0	195
pY	3	2	0	0	11

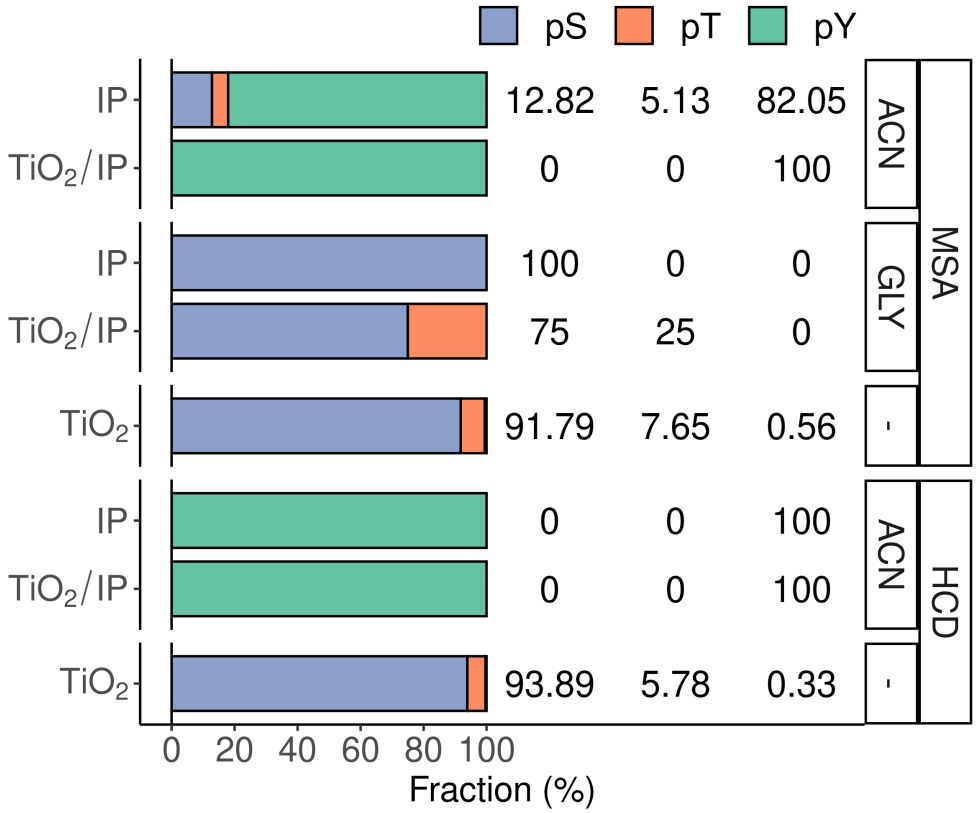

Figure 5: Fractions of phosphorylation sites discovered with each protocol variation. Visually presented as horizontal bar plots and numerically as percent values.

The different protocols yielded varying fractions of phosphorylation sites. Figure 5 shows the proportions of phosphorylated residues derived by each variation of the protocol. The regular phosphoproteomics protocol yields typical proportions of phosphorylated serines, threonines and tyrosines. This illustrates the issue of sub-stochiometric phosphotyrosine very well. The significant effect of immunoprecipitation for the enrichment of tyrosine-phosphorylated peptides can be clearly seen in the other protocols, provided acetonitrile/trifluoracetic acid has been used to elute peptides.

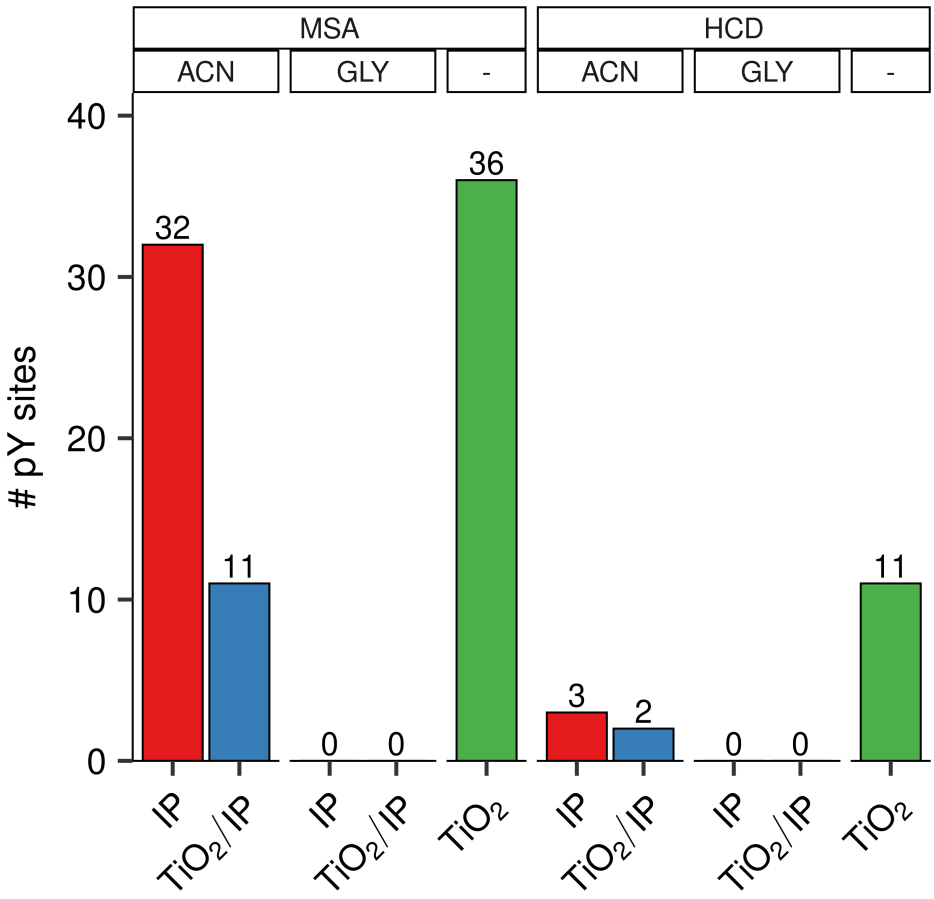

Figure 6: Phosphotyrosine sites discovered with each protocol variation. Number of phosphotyrosine sites with a phosphoRS probability score of $> 75\,\%$. Colors distinguish between protocol variations: red for IP, blue for TiO_2/IP and green for TiO_2.

Figure 6 shows the number of phosphotyrosine sites that could be discovered with a phosphoRS probability of greater than 75 %. Enrichment of phosphorylated peptides before the specific enrichment of phosphotyrosine-containing peptides reduced the number of discovered phosphotyrosine sites. However, not performing phosphotyrosine enrichment at all, i.e. relying upon the sensitivity of the MS instrument, gave rise to the most sites.

Low-pH glycine has shown to be not effective for this application. No phosphorylation sites could be recovered when using this elution method. The fragmentation method MSA outperformed HCD significantly, which was already seen in previous experiments (Section 4.1).

The overlaps of the identified phosphotyrosine sites in each protocol can be seen

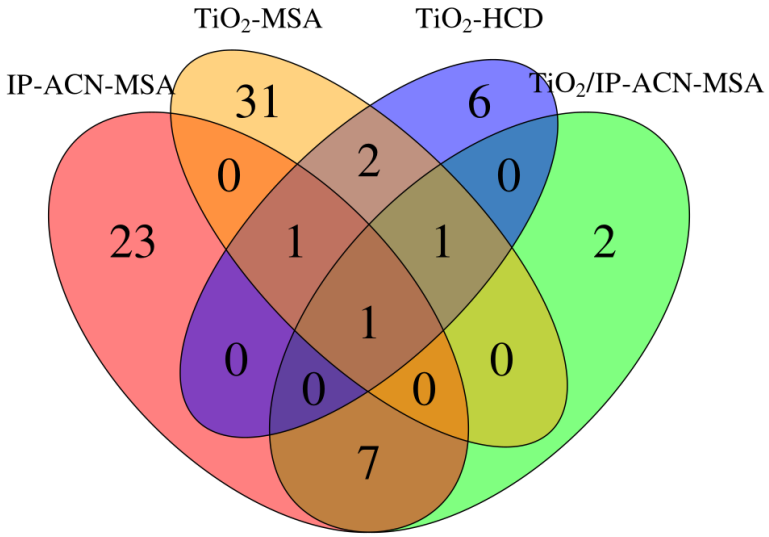

Figure 7: Venn diagram showing the number of phosphotyrosine identifications by each protocol.

in Figure 7. The majority of phosphotyrosine sites were identified in only one of the protocols. A notable overlap is between the protocols of IP and TiO_2/IP, both with ACN elution and MSA fragmentation, mutually identifying 8 sites of tyrosine phosphorylation. This might indicate that enrichment by immunoprecipitation could be a more reproducible approach than the protocol which only uses TiO_2 for enrichment of all phosphorylated peptides. Hence, albeit the TiO_2-MSA protocol revealed the most phosphotyrosine sites, specifically enriching for phosphotyrosine, by the means of immunoprecipitation, might be more suited in practical applications.

4.5 Phosphorylation site quality measures

In the computational analysis of raw mass spectrometry data for phosphoproteomics, phosphoRS is used to assess the probability of the correct assignment of a phosphorylation site. This algorithm is discussed in the introduction of this report in Section 1.4.5. To gain a deeper insight in the different methods that can be utilized for phosphotyrosine enrichment and analysis, the phosphoRS probability score is used as a measure for the quality of detections by each method. Here, the different variations of the protocol are compared in terms of this quality measure.

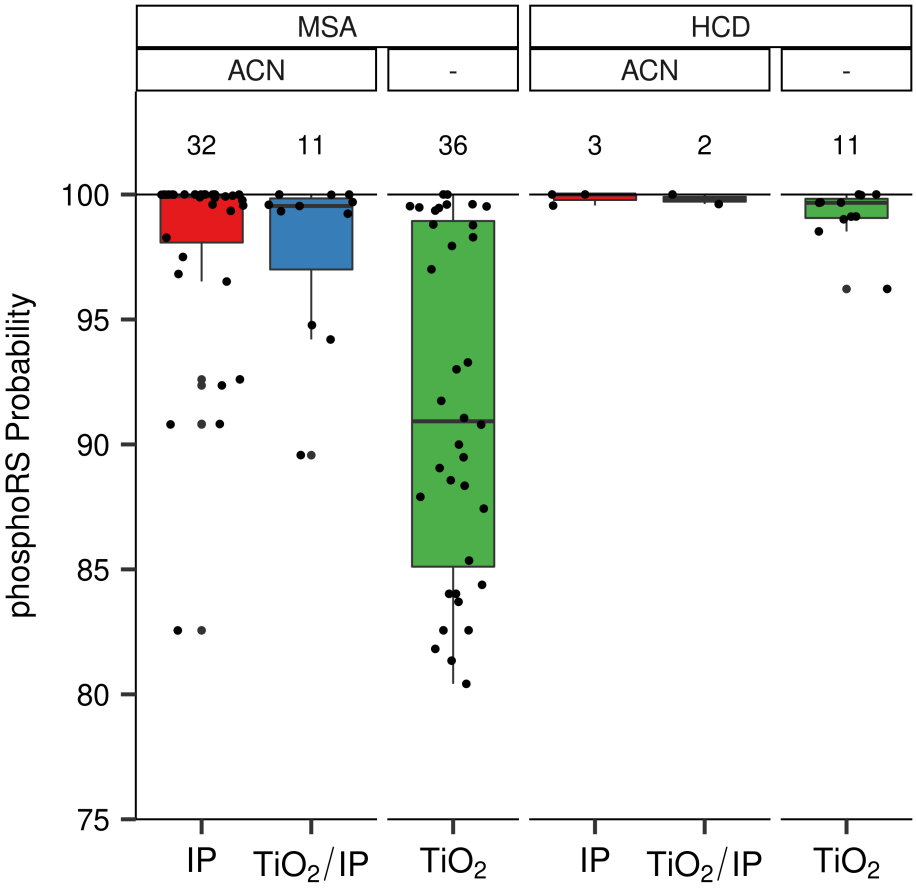

Figure 8: Distribution of phosphoRS Best Site Probabilities for each variation of the protocol. Above boxplots, the total number of phosphorylated tyrosine residues is shown. Colors distinguish between protocol variations: red for IP, blue for TiO_2/IP and green for TiO_2.

Figure 8 shows the distribution of the phosphoRS Best Site Probability scores among the different protocol variations, from all phosphotyrosine sites with a score of greater than 75 %. Even though immunoprecipitation enrichment of phosphotyrosine peptides

leads to fewer detectable phosphotyrosine sites, it achieves significantly better measures in terms of phosphoRS probability (Welch's two-sample t-test between IP-ACN-MSA and TiO_2-MSA: $p = 1.248 \times 10^{-5}$) and therefore higher quality detections than the general phosphopeptide enrichment (IP-ACN-MSA: 98.0 ± 3.8 %; TiO_2-MSA: 91.6 ± 6.8 %).

Table 3: Fractions of high phosphoRS probabilities in IP-ACN-MSA and TiO_2-MSA.

	IP-ACN-MSA		TiO_2-MSA	
phosphoRS probability	Count	Percent	Count	Percent
> 75 %	32	100	36	100
> 95 %	27	84.38	14	38.89
100 %	12	37.5	2	5.56

Table 3 shows the counts as well as percentages of phosphosites detected with phosphoRS probabilities > 75, > 95 or 100 % of both methods. In total, the IP-ACN-MSA method yielded more than a third (37.5 %) of its phosphotyrosine sites with a 100 % phosphoRS probability, while only a comparable fraction (38.89 %) of TiO_2-MSA were above 95 %.

With the protocol described in this report, immunoprecipitation yields fewer phosphotyrosine sites compared to TiO_2 enrichment. However, the overall quality of the detected sites is significantly enhanced by enriching specifically for phosphotyrosine peptides prior to LS-MS/MS.

4.6 Tyrosine-phosphorylated proteins

Proteins that were identified to be phosphorylated on tyrosine are shown in Table 5 and 6 for the protocols IP-ACN-MSA and TiO_2-MSA, respectively.

Among the proteins that were identified are mitogen-activated protein kinases (MAPK) which are key parts in many pathways. Characteristic for these proteins are TxY motifs in the activation loop of the kinase that contain threonine and tyrosine residues that need to be phosphorylated for the protein to reside in a catalytically active conformation [120]. Due to the implications of these kinases in the regulation of various cell functions like proliferation, gene expression, mitosis and apoptosis, they would be of high interest in any study of the proteome of cell lines for biopharmaceutical applications. Table 4 shows three identified MAP kinases and the peptide sequence with the characteristic motif and the phosphorylated tyrosine.

Table 4: MAP kinases identified with the IP-ACN-MSA protocol. Highlighted in red is the TxY motif. Phosphorylation of tyrosine is indicated with a preceding lowercase p.

	Peptide Sequence	Position in Protein
MAPK7	[R].GLCTSPAEHQYFM TEpY VATR.[W]	591–610
MAPK14	[R].HTDDEM TGpY VATR.[W]	163–175
MAPK3	[R].IADPEHDHTGFL TEpY VATR.[W]	219–237

Another protein that was identified as being tyrosine-phosphorylated is vimentin. With the IP protocol, using ACN elution and MSA fragmentation, the phosphotyrosine at amino acid position 117 was identified, while the TiO_2 protocols could only find phosphorylated serine residues. Vimentin is an intermediate filament protein and part of the cytoskeleton. It has been shown that vimentin is tightly regulated by phosphorylation [121]. Many studies have been conducted that have shown differential expression and alterations in phosphorylation of vimentin upon temperature shift in CHO cell cultures [39, 38, 122].

The Crk-like, identified with a tyrosine phosphorylation at position 207 with the protocol IP-ACN-MSA, is a protein that has a SH3 and a SH2 domain. The first is a common protein domain that recognizes proline-rich motifs; the latter recognizes phosphorylated tyrosine residues. CRKL is significantly involved in signal transduction

throughout various pathways, making it a key part in many biological processes such as the regulation of cell growth and gene expression.

The IP-ACN-MSA protocol further revealed tyrosine-phosphorylation of Signal Transducer and Activator of Transcription 3 (STAT3) at position 705. This particular phosphorylation site is known to be involved in signaling processes regulating differentiation, proliferation and apoptosis; most prominently via the JAK-STAT signaling pathway [123].

Table 5: Tyrosine-phosphorylated proteins identified with the IP-ACN-MSA protocol with the position of the phosphorylated tyrosine residue and the probability.

Protein names	Phosphorylated Postion	phosphoRS: Best Site Probability
Partitioning defective 3-like	Y 161	99.92
Non-specific serine/threonine protein kinase (EC 2.7.11.1)	Y 48	100
UPF0663 transmembrane protein C17orf28-like	Y 443	99.99
Receptor-type tyrosine-protein phosphatase alpha	Y 789	99.99
CRKL (Crk-like protein)	Y 207	96.82
Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8)	Y 68	100
Palladin	Y 322	99.35
Dual specificity tyrosine-phosphorylation-regulated kinase 1A	Y 321	100
Protein S100-A10	Y 25	98.27
VIM (Vimentin)	Y 117	100
Interferon-inducible protein	Y 27	100
Glycogen synthase kinase-3 alpha	Y 371	99.77
Tyrosine-protein kinase (EC 2.7.10.2)	Y 220	92.6
Signal transducer and activator of transcription	Y 705	100
Serine/threonine-protein kinase ICK	Y 159	100
Mitogen-activated protein kinase (EC 2.7.11.24)	Y 171	99.99
ATG101 (Autophagy-related protein 101)	Y 164	100
Thioredoxin reductase 1, cytoplasmic	Y 226	100
Hepatocyte growth factor receptor	Y 1075	90.82
Mitogen-activated protein kinase (EC 2.7.11.24)	Y 606	90.8
Activated CDC42 kinase 1	Y 518	92.36
Mitogen-activated protein kinase (EC 2.7.11.24)	Y 233	99.56
Calmodulin	Y 83	99.87
60S acidic ribosomal protein P0	Y 24	100
Homeodomain-interacting protein kinase 1	Y 352	99.6
Uncharacterized protein	Y 113	99.89
Pyruvate kinase (EC 2.7.1.40)	Y 24	96.52

Table 6: Tyrosine-phosphorylated proteins identified with the TiO_2-MSA protocol with the position of the phosphorylated tyrosine residue and the probability.

Protein names	Phosphorylated Position	phosphoRS: Best Site Probability
Histone-lysine N-methyltransferase MLL4	Y 127	83.7
Biorientation of chromosomes in cell division protein 1-like	Y 2607	99.6
Caveolin	Y 19	87.9
DNA-directed RNA polymerase I subunit RPA43	Y 105	93.01
Microtubule-associated protein 1B	Y 1243	84.02
Microtubule-associated protein 1B	Y 1700	82.56
BRISC complex subunit Abro1	Y 377	80.42
INCENP (Inner centromere protein)	Y 761	97.01
DNA helicase (EC 3.6.4.12)	Y 159	88.35
Transcriptional regulator ATRX	Y 233	99.36
A-kinase anchor protein 12	Y 279	89.99
Nucleus accumbens-associated protein 1	Y 475	84.02
Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1	Y 1315	97.94
Zinc finger CCCH-type antiviral protein 1	Y 573	99.53
Zinc finger CCCH domain-containing protein 13	Y 926	88.57
Tubulin-folding cofactor B	Y 63	98.8
Tubby-like protein	Y 104	84.38
PH domain leucine-rich repeat-containing protein phosphatase 2	Y 1235	99.61
Putative E3 ubiquitin-protein ligase TRIP12	Y 1091	98.77
Tyrosine-protein kinase (EC 2.7.10.2)	Y 295	99.48
Glycogen synthase kinase-3 alpha	Y 371	91.06
Band 4.1-like protein 2	Y 88	93.28
RNA-binding protein 39	Y 95	99.52
Dual specificity tyrosine-phosphorylation-regulated kinase 1A	Y 321	100
Mitogen-activated protein kinase kinase kinase kinase 4	Y 254	81.35
Splicing factor	Y 237	90.79
Aryl-hydrocarbon-interacting protein-like 1	Y 236	100
LisH domain and HEAT repeat-containing protein KIAA1468	Y 192	91.74
Zinc finger transcription factor Trps1 (Fragment)	Y 194	87.43
Rap guanine nucleotide exchange factor 5	Y 69	89.49
Uncharacterized protein	Y 16	82.56
Gap junction protein	Y 286	81.82

4.7 Gene ontology analysis

Biological process and molecular function annotation terms of the found proteins are shown in Table 7. Gene Ontology terms were considered significant with a Benjamini-Hochberg adjusted $p < 0.05$ [111].

Among the gene ontology terms associated with tyrosine-phosphorylated proteins are biological processes of phosphorylation, and more specifically autophosphorylation which suggests that cell surface receptors involved in signal transduction were found.

Molecular function terms such as kinase activity and ATP binding further support this impression. Additionally to membrane-spanning cell-surface receptors, non-receptor tyrosine kinases might also be among the proteins, indicated by the molecular function term non-membrane spanning protein tyrosine kinase activity.

Notably, no Gene Ontology terms were found to be significantly enriched using any other protocol. The protocol IP-ACN-MSA therefore produced the most biologically meaningful results from all procedures that were compared.

Table 7: Gene ontology terms for biological process and molecular function derived by the IP-ACN-MSA method with the Benjamini-Hochberg adjusted p-value

Biological process		
GO ID	GO Term	adjusted p
GO:0016310	phosphorylation	3.43E-05
GO:0006468	protein phosphorylation	0.006333
GO:0038083	peptidyl-tyrosine autophosphorylation	0.027658
GO:0046777	protein autophosphorylation	0.028095
Molecular function		
GO ID	GO Term	adjusted p
GO:0016301	kinase activity	5.46E-04
GO:0004672	protein kinase activity	0.001490
GO:0004713	protein tyrosine kinase activity	0.002723
GO:0004715	non-membrane spanning protein tyrosine kinase activity	0.010541
GO:0016740	transferase activity	0.008612
GO:0005524	ATP binding	0.008147
GO:0019903	protein phosphatase binding	0.022713
GO:0000166	nucleotide binding	0.022812
GO:0004674	protein serine/threonine kinase activity	0.034846
GO:0005515	protein binding	0.048841

5 Discussion

The transient nature of tyrosine phosphorylation results in sub-stoichiometric amounts of tyrosine-phosphorylated proteins at any given moment in the cell. Sophisticated experimental techniques are necessary to gain insight into some of the crucial intracellular processes that rely on such modifications.

The most employed method for enrichment of phosphotyrosine-containing peptides is immunoprecipitation with anti-phosphotyrosine antibodies. Initial experiments were carried out to get an impression of the performance of the two available antibodies 4G10 and P-Tyr-1000. Under these circumstances, both antibodies failed to enrich a considerable number of phosphotyrosine sites. After deviating from manufacturer's recommendations and using higher concentrations of the antibody P-Tyr-1000, the enrichment of phosphotyrosine could eventually be observed. For the remainder of the project, this antibody has been used since it performed superior to the antibody 4G10.

For proteomic studies it is necessary to identify as much of the cell's proteins as possible. While MS instruments are getting increasingly sensitive, the sample preparation and processing steps that are taking place beforehand still cause dramatic loss of peptide and ultimately impede a thorough analysis. One step in the procedure of a phosphotyrosine enrichment workflow is the depletion of antibodies from the sample after immunoprecipitation, which is usually done using C18 spin columns. A study from Abe et al. [113] found significantly enhanced recovery of peptides when using Fe^{3+}-IMAC instead. Examinations of peptide recovery by TiO_2-based MOAC, a method similar to Fe^{3+}-IMAC, performed poorly compared to C18 spin columns based on measurements of peptide concentration. TiO_2 enrichment lead to a loss of 63.55 % of peptide, while the loss with C18 columns was only 33.26 %.

Since anti-phosphotyrosine antibodies have been reported to be sensitive towards sequence context, the identified phosphotyrosine-containing peptides were visualized in a sequence logo to investigate any striking biases in antibody binding. The antibody P-Tyr-1000 showed no major sensitivity towards sequence context throughout all analysed peptide sequences.

Literature research revealed a variety of approaches for enrichment of phosphoty-

rosine. Many previous studies were conducted with stimulation of epidermal growth factor receptor (EGFR) by its corresponding growth factor. This leads to elevated levels of tyrosine phosphorylation and therefore does not portray a natural state of the cells at the point of lysis. The scope of this thesis was to analyze tyrosine phosphorylation of biopharmaceutically relevant cell lines in a realistic, undistorted setting. It is therefore crucial to develop a procedure that is capable of enriching as much phosphotyrosine-containing peptides as possible. Modifications at several steps of the protocol, based on approaches found in the literature, gave rise to 10 different variations showing significantly different results with respect to the quantity and quality of identified phosphotyrosine sites.

Among the variations of the protocol is the optional pre-enrichment of global phosphopeptides before the specific phosphotyrosine enrichment. While pre-enrichment is a common step in a phosphotyrosine-enrichment procedure, any additional preparation step causes more loss of sample, hence sample handling should be kept at a minimum. It was therefore investigated which potential benefit arises from omitting the global phosphopeptide enrichment. The protocol which used just immunoprecipitation from whole peptide outperformed the protocol with pre-enrichment in terms of the number of discovered phosphotyrosine sites. The protocol IP-ACN-MSA gave rise to 32 sites, while the TiO_2/IP-ACN-MSA protocol only discovered 11. Furthermore, 8 sites were mutually identified with both protocols which might indicate a high reproducibility of using immunoprecipitation.

Advancements in mass spectrometry technology have led to improvements in depth and resolution of proteomics analyses. At some point, these improvements might outweigh the benefit of — and the need for — specific enrichment of sub-stoichiometrically occurring phosphotyrosine. Therefore, additionally to protocols with immunoprecipitation, a regular phosphoproteomics protocol was conducted to compare the number and quality of discovered phosphotyrosine sites. The distribution of phosphorylated serine, threonine and tyrosine sites illustrates the rarity of tyrosine phosphorylation very well. Using MSA fragmentation, 5948 phosphoserine and 496 phosphothreonine sites were discovered, however, only 36 phosphotyrosines out of a total of 6480 phosphorylated sites were identified. Quantitatively, however, this still outnumbers the best performing protocol using immunoprecipitation, which yielded 32 sites. A striking difference

between the protocols is shown by the phosphoRS probability, a measure of confidence for the assignation of a phosphorylation at a tyrosine site. Immunoprecipitation-based protocols yield phosphotyrosine sites with very high probability, with 12 out of 32 phosphotyrosine sites having a probability of 100 %, compared to the phosphoproteomics procedure with only 2 phosphotyrosine sites out of 36.

Fragmentation of peptides plays a crucial role in the discovery of phosphorylated sites. Two fragmentation methods that are usually employed for phosphoproteomics experiments are high-energy C-trap dissociation (HCD) and multistage activation (MSA). To elucidate which of these methods perform better for the underlying task, each sample has been subjected to two LS-MS/MS runs, one for each fragmentation method. Throughout all experiments it was shown that MSA performs superiorly overall. MSA fragmented results yielded more general peptide as well as more phosphorylated peptides and sites for global phosphorylations as well as tyrosine specifically.

Proteins that were discovered with the developed protocol might be of interest in studies of the tyrosine proteome. Among them are mitogen-activated protein kinases (MAPK), Vimentin (VIM), Crk-like protein (CRKL) and signal transducer and activator of transcription 3 (STAT3). For MAPK, characteristic phosphorylation motifs were identified that yield information about the catalytic activity of the protein [120]. VIM is a protein that has been subject of investigations regarding CHO cell culturing with temperature shift [39] and STAT3 is a key member of important intracellular signaling pathways [123]. Gene Ontology Term Enrichment Analysis (GOTEA) has revealed biological processes of (auto)phosphorylation as well as molecular functions involving kinase activities, ATP binding and nucleotide binding among others.

5.1 Conclusions & Outlook

Specific enrichment of phosphotyrosine has been shown to improve the confidence of phosphorylated site inference algorithms. Furthermore, preliminary results show that the reproducibility might be improved by immunoprecipitation methods. The most favorable protocol was found to be antibody-based immunoaffinity enrichment of tyrosine-phosphorylated peptides from total peptide, without pre-enrichment of phosphorylated peptides by TiO_2. For the elution method it is recommended to use 60 % acetoni-

trile and 0.1 % trifluoracetic acid as opposed to low-pH glycine. For fragmentation, multistage activation (MSA) has been demonstrated to be best suited.

While omitting the specific enrichment of phosphotyrosine yielded the most phosphotyrosine residues, this probability of phosphorylated sites was decreased significantly and no Gene Ontology terms were found significantly enriched.

Future investigations should focus on further improving the yield in terms of quantity of phosphotyrosine sites. This might be achieved by drastically increasing the amount of peptide for the protocol. It should be noted that the results in this thesis were obtained from 1 mg of protein per run while previous studies generally used much larger quantities. In terms of quality, the proposed protocol already shows promising results.

Studies of phopshorylation events in CHO cells have been successfully applied to aid the development of industrially relevant phenotypes and to enhance bioproduction efficacy. Comprehensive profiling of tyrosine phosphorylation allows for an even broader insight into critical cellular processes and facilitates the rational approaches to cellular and process engineering.

References

[1] Y. L. Deribe, T. Pawson, and I. Dikic, "Post-translational modifications in signal integration," *Nature Structural & Molecular Biology*, vol. 17, pp. 666–672, June 2010.

[2] N. Jenkins, R. Parekh, and D. C. James, "Getting the glycosylation right: Implications for the biotechnology industry," *Nature Biotechnology*, Aug. 1996.

[3] T. Omasa, M. Onitsuka, and W.-D. Kim, "Cell Engineering and Cultivation of Chinese Hamster Ovary (CHO) Cells," *Current Pharmaceutical Biotechnology*, vol. 11, pp. 233–240, Apr. 2010.

[4] G. A. Khoury, R. C. Baliban, and C. A. Floudas, "Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database," *Scientific Reports*, vol. 1, Sept. 2011.

[5] P. Minguez, L. Parca, F. Diella, D. R. Mende, R. Kumar, M. Helmer-Citterich, A.-C. Gavin, V. van Noort, and P. Bork, "Deciphering a global network of functionally associated post-translational modifications," *Molecular Systems Biology*, vol. 8, p. 599, July 2012.

[6] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam, "The protein kinase complement of the human genome," *Science (New York, N.Y.)*, vol. 298, pp. 1912–1934, June 2002.

[7] S. Caenepeel, G. Charydczak, S. Sudarsanam, T. Hunter, and G. Manning, "The mouse kinome: Discovery and comparative genomics of all mouse protein kinases," *Proceedings of the National Academy of Sciences*, vol. 101, pp. 11707–11712, Aug. 2004.

[8] A. S. Gajadhar and F. M. White, "System level dynamics of post-translational modifications," *Current opinion in biotechnology*, vol. 0, pp. 83–87, Aug. 2014.

[9] E. J. Needham, B. L. Parker, T. Burykin, D. E. James, and S. J. Humphrey, "Illuminating the dark phosphoproteome," *Science Signaling*, vol. 12, p. eaau8645, Jan. 2019.

[10] T. Hunter, "THE CROONIAN LECTURE 1997. The phosphorylation of proteins on tyrosine: Its role in cell growth and disease," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 353, pp. 583–605, Apr. 1998.

[11] G. Manning, "The Protein Kinase Complement of the Human Genome," *Science*, vol. 298, pp. 1912–1934, Dec. 2002.

[12] W. Eckhart, M. A. Hutchinson, and T. Hunter, "An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates," *Cell*, vol. 18, pp. 925–933, Dec. 1979.

[13] T. Hunter, "The Genesis of Tyrosine Phosphorylation," *Cold Spring Harbor Perspectives in Biology*, vol. 6, May 2014.

[14] K. Sharma, R. C. J. D'Souza, S. Tyanova, C. Schaab, J. R. Wiśniewski, J. Cox, and M. Mann, "Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling," *Cell Reports*, vol. 8, pp. 1583–1594, Sept. 2014.

[15] T. Hunter, "Tyrosine phosphorylation: Thirty years and counting," *Current Opinion in Cell Biology*, vol. 21, pp. 140–146, Apr. 2009.

[16] M. A. Lemmon and J. Schlessinger, "Cell signaling by receptor-tyrosine kinases," *Cell*, vol. 141, pp. 1117–1134, June 2010.

[17] G. Walsh, "Biopharmaceutical benchmarks 2018," *Nature Biotechnology*, vol. 36, pp. 1136–1145, Dec. 2018.

[18] J. H. Tjio and T. T. Puck, "Genetics of somatic mammalian cells. II. Chromosomal constitution of cells in tissue culture," *The Journal of Experimental Medicine*, vol. 108, pp. 259–268, Aug. 1958.

[19] G. Ringold, B. Dieckmann, and F. Lee, "Co-expression and amplification of dihydrofolate reductase cDNA and the Escherichia coli XGPRT gene in Chinese hamster ovary cells," *Journal of Molecular and Applied Genetics*, vol. 1, no. 3, pp. 165–175, 1981.

[20] R. J. Kaufman and P. A. Sharp, "Amplification and expression of sequences

cotransfected with a modular dihydrofolate reductase complementary dna gene," *Journal of Molecular Biology*, vol. 159, pp. 601–621, Aug. 1982.

[21] S. J. Scahill, R. Devos, J. Van der Heyden, and W. Fiers, "Expression and characterization of the product of a human immune interferon cDNA gene in Chinese hamster ovary cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 80, pp. 4654–4658, Aug. 1983.

[22] D. L. Hacker, M. De Jesus, and F. M. Wurm, "25 years of recombinant proteins from reactor-grown cells — Where do we go from here?," *Biotechnology Advances*, vol. 27, pp. 1023–1027, Nov. 2009.

[23] M. Matasci, D. L. Hacker, L. Baldi, and F. M. Wurm, "Recombinant therapeutic protein production in cultivated mammalian cells: Current status and future prospects," *Drug Discovery Today: Technologies*, vol. 5, pp. e37–e42, Sept. 2008.

[24] J. R. Birch and R. Arathoon, "Suspension culture of mammalian cells," *Bioprocess Technology*, vol. 10, pp. 251–270, 1990.

[25] F. M. Wurm, "Production of recombinant protein therapeutics in cultivated mammalian cells," *Nature Biotechnology*, vol. 22, pp. 1393–1398, Nov. 2004.

[26] J. S. Lee and G. M. Lee, "Rapamycin treatment inhibits CHO cell death in a serum-free suspension culture by autophagy induction," *Biotechnology and Bioengineering*, vol. 109, pp. 3093–3102, Dec. 2012.

[27] Y. Kim, E. Baek, J. Lee, and G. Lee, "Autophagy and its implication in Chinese hamster ovary cell culture," *Biotechnology letters*, vol. 35, July 2013.

[28] B. S. Majors, M. J. Betenbaugh, N. E. Pederson, and G. G. Chiang, "Mcl-1 overexpression leads to higher viabilities and increased production of humanized monoclonal antibody in Chinese hamster ovary cells," *Biotechnology Progress*, vol. 25, no. 4, pp. 1161–1168, 2009 Jul-Aug.

[29] M. Zhou, Y. Crawford, D. Ng, J. Tung, A. F. J. Pynn, A. Meier, I. H. Yuk, N. Vijayasankaran, K. Leach, J. Joly, B. Snedecor, and A. Shen, "Decreasing lactate level and increasing antibody production in Chinese Hamster Ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase

kinases," *Journal of Biotechnology*, vol. 153, pp. 27–34, Apr. 2011.

[30] M. Jeon, D. Yu, and D. Lee, "Combinatorial engineering of ldh-a and bcl-2 for reducing lactate production and improving cell growth in dihydrofolate reductase-deficient Chinese hamster ovary cells.," *Applied Microbiology and Biotechnology*, vol. 92, pp. 779–790, July 2011.

[31] S. Fischer, R. Handrick, and K. Otte, "The art of CHO cell engineering: A comprehensive retrospect and future perspectives," *Biotechnology Advances*, vol. 33, pp. 1878–1896, Dec. 2015.

[32] N. Barron, N. Kumar, N. Sanchez, P. Doolan, C. Clarke, P. Meleady, F. O'Sullivan, and M. Clynes, "Engineering CHO cell growth and recombinant protein productivity by overexpression of miR-7," *Journal of Biotechnology*, vol. 151, pp. 204–211, Jan. 2011.

[33] S. Fischer, K. F. Marquart, L. A. Pieper, J. Fieder, M. Gamer, I. Gorr, P. Schulz, and H. Bradl, "miRNA engineering of CHO cells facilitates production of difficult-to-express proteins and increases success in cell line development," *Biotechnology and Bioengineering*, vol. 114, pp. 1495–1510, July 2017.

[34] O. Coleman, S. Suda, J. Meiller, M. Henry, M. Riedl, N. Barron, M. Clynes, and P. Meleady, "Increased growth rate and productivity following stable depletion of miR-7 in a mAb producing CHO cell line causes an increase in proteins associated with the Akt pathway and ribosome biogenesis," *Journal of Proteomics*, vol. 195, pp. 23–32, Mar. 2019.

[35] H. Kaufmann, X. Mazur, M. Fussenegger, and J. E. Bailey, "Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells," *Biotechnology and Bioengineering*, vol. 63, no. 5, pp. 573–582, 1999.

[36] E. Trummer, K. Fauland, S. Seidinger, K. Schriebl, C. Lattenmayer, R. Kunert, K. Vorauer-Uhl, R. Weik, N. Borth, H. Katinger, and D. Müller, "Process parameter shifting: Part II. Biphasic cultivation-A tool for enhancing the volumetric productivity of batch processes using Epo-Fc expressing CHO cells," *Biotechnology and Bioengineering*, vol. 94, pp. 1045–1052, Aug. 2006.

[37] M. Yandell and D. Ence, "A beginner's guide to eukaryotic genome annotation,"

Nature Reviews Genetics, vol. 13, pp. 329–342, May 2012.

[38] N. Kumar, P. Gammell, P. Meleady, M. Henry, and M. Clynes, "Differential protein expression following low temperature culture of suspension CHO-K1 cells," *BMC Biotechnology*, vol. 8, p. 42, Apr. 2008.

[39] M. Henry, M. Power, P. Kaushik, O. Coleman, M. Clynes, and P. Meleady, "Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift," *Journal of Proteome Research*, vol. 16, pp. 2339–2358, July 2017.

[40] C. Y. X. Chan, M. A. Gritsenko, R. D. Smith, and W.-J. Qian, "The current state of the art of quantitative phosphoproteomics and its applications to diabetes research," *Expert review of proteomics*, vol. 13, pp. 421–433, Apr. 2016.

[41] H. Zhou, M. Ye, J. Dong, G. Han, X. Jiang, R. Wu, and H. Zou, "Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis," *Journal of Proteome Research*, vol. 7, pp. 3957–3967, Sept. 2008.

[42] M. C. Posewitz and P. Tempst, "Immobilized gallium(III) affinity chromatography of phosphopeptides," *Analytical Chemistry*, vol. 71, pp. 2883–2892, July 1999.

[43] S. Feng, M. Ye, H. Zhou, X. Jiang, X. Jiang, H. Zou, and B. Gong, "Immobilized Zirconium Ion Affinity Chromatography for Specific Enrichment of Phosphopeptides in Phosphoproteome Analysis," *Molecular & Cellular Proteomics*, vol. 6, pp. 1656–1665, Sept. 2007.

[44] D. C. Neville, C. R. Rozanas, E. M. Price, D. B. Gruis, A. S. Verkman, and R. R. Townsend, "Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry," *Protein Science: A Publication of the Protein Society*, vol. 6, pp. 2436–2445, Nov. 1997.

[45] S. B. Ficarro, M. L. McCleland, P. T. Stukenberg, D. J. Burke, M. M. Ross, J. Shabanowitz, D. F. Hunt, and F. M. White, "Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae," *Nature Biotechnology*, vol. 20, pp. 301–305, Mar. 2002.

[46] M. R. Larsen, T. E. Thingholm, O. N. Jensen, P. Roepstorff, and T. J. D. Jørgensen, "Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns," *Molecular & Cellular Proteomics*, vol. 4, pp. 873–886, July 2005.

[47] S. S. Jensen and M. R. Larsen, "Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques," *Rapid Communications in Mass Spectrometry*, vol. 21, pp. 3635–3645, Nov. 2007.

[48] A. Leitner, "Phosphopeptide enrichment using metal oxide affinity chromatography," *TrAC Trends in Analytical Chemistry*, vol. 29, pp. 177–185, Feb. 2010.

[49] M. R. Larsen, T. E. Thingholm, O. N. Jensen, P. Roepstorff, and T. J. D. Jørgensen, "Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns," *Molecular & cellular proteomics: MCP*, vol. 4, pp. 873–886, July 2005.

[50] T. E. Thingholm, O. N. Jensen, P. J. Robinson, and M. R. Larsen, "SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides," *Molecular & cellular proteomics: MCP*, vol. 7, pp. 661–671, Apr. 2008.

[51] P. J. Boersema, L. Y. Foong, V. M. Y. Ding, S. Lemeer, B. van Breukelen, R. Philp, J. Boekhorst, B. Snel, J. den Hertog, A. B. H. Choo, and A. J. R. Heck, "In-depth Qualitative and Quantitative Profiling of Tyrosine Phosphorylation Using a Combination of Phosphopeptide Immunoaffinity Purification and Stable Isotope Dimethyl Labeling," *Molecular & Cellular Proteomics : MCP*, vol. 9, pp. 84–99, Jan. 2010.

[52] J. Rush, A. Moritz, K. A. Lee, A. Guo, V. L. Goss, E. J. Spek, H. Zhang, X.-M. Zha, R. D. Polakiewicz, and M. J. Comb, "Immunoaffinity profiling of tyrosine phosphorylation in cancer cells," *Nature Biotechnology*, vol. 23, pp. 94–101, Jan. 2005.

[53] Y. Zhang, A. Wolf-Yadlin, P. L. Ross, D. J. Pappin, J. Rush, D. A. Lauffenburger, and F. M. White, "Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals

Dynamic Modules," *Molecular & Cellular Proteomics*, vol. 4, pp. 1240–1250, Sept. 2005.

[54] S. G. Ward, K. Reif, S. Ley, M. J. Fry, M. D. Waterfield, and D. A. Cantrell, "Regulation of phosphoinositide kinases in T cells. Evidence that phosphatidylinositol 3-kinase is not a substrate for T cell antigen receptor-regulated tyrosine kinases," *The Journal of Biological Chemistry*, vol. 267, pp. 23862–23869, Nov. 1992.

[55] J. R. Glenney, L. Zokas, and M. P. Kamps, "Monoclonal antibodies to phosphotyrosine," *Journal of Immunological Methods*, vol. 109, pp. 277–285, May 1988.

[56] Y. Bian, L. Li, M. Dong, X. Liu, T. Kaneko, K. Cheng, H. Liu, C. Voss, X. Cao, Y. Wang, D. Litchfield, M. Ye, S. S.-C. Li, and H. Zou, "Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder," *Nature Chemical Biology*, vol. 12, pp. 959–966, Nov. 2016.

[57] A. N. Kettenbach and S. A. Gerber, "Rapid and Reproducible Single-Stage Phosphopeptide Enrichment of Complex Peptide Mixtures: Application to General and Phosphotyrosine-Specific Phosphoproteomics Experiments," *Analytical Chemistry*, vol. 83, pp. 7635–7644, Oct. 2011.

[58] K. Sharma, R. C. J. D'Souza, S. Tyanova, C. Schaab, J. R. Wiśniewski, J. Cox, and M. Mann, "Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling," *Cell Reports*, vol. 8, pp. 1583–1594, Sept. 2014.

[59] B. Blagoev, S.-E. Ong, I. Kratchmarova, and M. Mann, "Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics," *Nature Biotechnology*, vol. 22, pp. 1139–1145, Sept. 2004.

[60] J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, and M. Mann, "Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks," *Cell*, vol. 127, pp. 635–648, Nov. 2006.

[61] X. Han, A. Aslanian, and J. R. Yates, "Mass Spectrometry for Proteomics," *Current opinion in chemical biology*, vol. 12, pp. 483–490, Oct. 2008.

[62] E. Sabidó, N. Selevsek, and R. Aebersold, "Mass spectrometry-based proteomics for systems biology," *Current Opinion in Biotechnology*, vol. 23, pp. 591–597, Aug. 2012.

[63] M. Larsen and A. Edwards, "Phosphoproteomics: Approaches, Developments, and Challenges," in *Encyclopedia of Cell Biology*, pp. 177–186, Elsevier, 2016.

[64] A. El-Aneed, A. Cohen, and J. Banoub, "Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers," *Applied Spectroscopy Reviews*, vol. 44, pp. 210–230, Apr. 2009.

[65] C. Ho, C. Lam, M. Chan, R. Cheung, L. Law, L. Lit, K. Ng, M. Suen, and H. Tai, "Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications," *The Clinical Biochemist Reviews*, vol. 24, pp. 3–12, Feb. 2003.

[66] N. Singhal, M. Kumar, P. K. Kanaujia, and J. S. Virdi, "MALDI-TOF mass spectrometry: An emerging technology for microbial identification and diagnosis," *Frontiers in Microbiology*, vol. 6, 2015.

[67] A. Cohen, A. Mansour, and J. Banoub, "Absolute quantification of Atlantic salmon and rainbow trout vitellogenin by the 'signature peptide' approach using electrospray ionization QqToF tandem mass spectrometry." https://pubmed.ncbi.nlm.nih.gov/16541402/, May 2006.

[68] G. Grant, S. Frison, J. Yeung, T. Vasanthan, and P. Sporns, "Comparison of MALDI-TOF mass spectrometric to enzyme colorimetric quantification of glucose from enzyme-hydrolyzed starch." https://pubmed.ncbi.nlm.nih.gov/14518935/, Aug. 2003.

[69] K. Y. C. Fung, S. Askovic, F. Basile, and M. W. Duncan, "A simple and inexpensive approach to interfacing high-performance liquid chromatography and matrix-assisted laser desorption/ionization-time of flight-mass spectrometry," *PROTEOMICS*, vol. 4, no. 10, pp. 3121–3127, 2004.

[70] M. W. Duncan, H. Roder, and S. W. Hunsucker, "Quantitative matrix-assisted laser desorption/ionization mass spectrometry," *Briefings in Functional Genomics and Proteomics*, vol. 7, p. 355, Sept. 2008.

[71] J. H. McReynolds and M. Anbar, "Isotopic assay of nanomole amounts of nitrogen-15 labeled amino acids by collision-induced dissociation mass spectrometry," *Analytical Chemistry*, vol. 49, pp. 1832–1836, Oct. 1977.

[72] P. J. Boersema, S. Mohammed, and A. J. R. Heck, "Phosphopeptide fragmentation and analysis by mass spectrometry," *Journal of Mass Spectrometry*, vol. 44, pp. 861–878, June 2009.

[73] P. Roepstorff and J. Fohlman, "Proposal for a common nomenclature for sequence ions in mass spectra of peptides.," *Biomedical Mass Spectrometry*, vol. 11, pp. 601–601, Nov. 1984.

[74] V. H. Wysocki, G. Tsaprailis, L. L. Smith, and L. A. Breci, "Mobile and localized protons: A framework for understanding peptide dissociation," *J. Mass Spectrom.*, p. 8, 2000.

[75] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, pp. 44–57, Jan. 2009.

[76] Y. Huang, G. C. Tseng, S. Yuan, L. Pasa-Tolic, M. S. Lipton, R. D. Smith, and V. H. Wysocki, "A Data-Mining Scheme for Identifying Peptide Structural Motifs Responsible for Different MS/MS Fragmentation Intensity Patterns," *Journal of proteome research*, vol. 7, pp. 70–79, Jan. 2008.

[77] A. M. Palumbo, S. A. Smith, C. L. Kalcic, M. Dantus, P. M. Stemmer, and G. E. Reid, "Tandem mass spectrometry strategies for phosphoproteome analysis: MS/MS STRATEGIES FOR PHOSPHOPROTEOME ANALYSIS," *Mass Spectrometry Reviews*, vol. 30, pp. 600–625, July 2011.

[78] R. Brown, S. A. Stuart, S. Houel, N. G. Ahn, and W. M. Old, "Large-Scale Examination of Factors Influencing Phosphopeptide Neutral Loss during Collision Induced Dissociation," *Journal of the American Society for Mass Spectrometry*, vol. 26, pp. 1128–1142, July 2015.

[79] R. A. Everley, E. L. Huttlin, A. R. Erickson, S. A. Beausoleil, and S. P. Gygi, "Neutral Loss is a Very Common Occurrence in Phosphotyrosine-containing Peptides Labeled with Isobaric Tags," *Journal of proteome research*, vol. 16, pp. 1069–

1076, Feb. 2017.

[80] M. J. Schroeder, J. Shabanowitz, J. C. Schwartz, D. F. Hunt, and J. J. Coon, "A Neutral Loss Activation Method for Improved Phosphopeptide Sequence Analysis by Quadrupole Ion Trap Mass Spectrometry," *Analytical Chemistry*, vol. 76, pp. 3590–3598, July 2004.

[81] J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, and M. Mann, "Higher-energy C-trap dissociation for peptide modification analysis," *Nature Methods*, vol. 4, pp. 709–712, Sept. 2007.

[82] N. Nagaraj, R. C. J. D'Souza, J. Cox, J. V. Olsen, and M. Mann, "Correction to Feasibility of Large-Scale Phosphoproteomics with Higher Energy Collisional Dissociation Fragmentation," *Journal of Proteome Research*, vol. 11, pp. 3506–3508, June 2012.

[83] M. P. Jedrychowski, E. L. Huttlin, W. Haas, M. E. Sowa, R. Rad, and S. P. Gygi, "Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics," *Molecular & cellular proteomics: MCP*, vol. 10, p. M111.009910, Dec. 2011.

[84] C. K. Frese, A. F. M. Altelaar, M. L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A. J. R. Heck, and S. Mohammed, "Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos," *Journal of Proteome Research*, vol. 10, pp. 2377–2388, May 2011.

[85] R. A. Zubarev, N. L. Kelleher, and F. W. McLafferty, "Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process," *Journal of the American Chemical Society*, vol. 120, pp. 3265–3266, Apr. 1998.

[86] E. A. Syrstad and F. Tureček, "Toward a general mechanism of electron capture dissociation," *Journal of the American Society for Mass Spectrometry*, vol. 16, pp. 208–224, Feb. 2005.

[87] S. M. M. Sweet, C. M. Bailey, D. L. Cunningham, J. K. Heath, and H. J. Cooper, "Large Scale Localization of Protein Phosphorylation by Use of Electron Capture Dissociation Mass Spectrometry," *Molecular & Cellular Proteomics*, vol. 8, pp. 904–912, May 2009.

[88] J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, and D. F. Hunt, "Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 9528–9533, June 2004.

[89] D. J. Douglas, A. J. Frank, and D. Mao, "Linear ion traps in mass spectrometry," *Mass Spectrometry Reviews*, vol. 24, pp. 1–29, Jan. 2005.

[90] A. D. Catherman, O. S. Skinner, and N. L. Kelleher, "Top Down Proteomics: Facts and Perspectives," *Biochemical and biophysical research communications*, vol. 445, pp. 683–693, Mar. 2014.

[91] A. Michalski, E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Müller, R. Viner, J. Schwartz, P. Remes, M. Belford, J.-J. Dunyach, J. Cox, S. Horning, M. Mann, and A. Makarov, "Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes," *Molecular & cellular proteomics: MCP*, vol. 11, p. O111.013698, Mar. 2012.

[92] G. C. McAlister, W. T. Berggren, S. Horning, A. Makarov, D. Phanstiel, J. Griep-Raming, G. Stafford, D. L. Swaney, J. E. P. Syka, V. Zabrouskov, and J. J. Coon, "A Proteomics Grade Electron Transfer Dissociation-enabled Hybrid Linear Ion Trap-orbitrap Mass Spectrometer," *Journal of proteome research*, vol. 7, pp. 3127–3136, Aug. 2008.

[93] A. Michalski, E. Damoc, J.-P. Hauschild, O. Lange, A. Wieghaus, A. Makarov, N. Nagaraj, J. Cox, M. Mann, and S. Horning, "Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer," *Molecular & cellular proteomics: MCP*, vol. 10, p. M111.011015, Sept. 2011.

[94] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, and B. Kuster, "Quantitative mass spectrometry in proteomics: A critical review," *Analytical and Bioanalytical Chemistry*, vol. 389, pp. 1017–1031, Oct. 2007.

[95] J. Li, J. G. Van Vranken, L. Pontano Vaites, D. K. Schweppe, E. L. Huttlin, C. Etienne, P. Nandhikonda, R. Viner, A. M. Robitaille, A. H. Thompson, K. Kuhn,

I. Pike, R. D. Bomgarden, J. C. Rogers, S. P. Gygi, and J. A. Paulo, "TMTpro reagents: A set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples," *Nature Methods*, vol. 17, pp. 399–404, Apr. 2020.

[96] A. Thompson, N. Wölmer, S. Koncarevic, S. Selzer, G. Böhm, H. Legner, P. Schmid, S. Kienle, P. Penning, C. Höhle, A. Berfelde, R. Martinez-Pinna, V. Farztdinov, S. Jung, K. Kuhn, and I. Pike, "TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set," *Analytical Chemistry*, vol. 91, pp. 15941–15950, Dec. 2019.

[97] Y. Zhang, B. R. Fonslow, B. Shan, M.-C. Baek, and J. R. Yates, "Protein Analysis by Shotgun/Bottom-up Proteomics," *Chemical reviews*, vol. 113, pp. 2343–2394, Apr. 2013.

[98] N. M. Riley and J. J. Coon, "Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling," *Analytical Chemistry*, vol. 88, pp. 74–94, Jan. 2016.

[99] K. A. Neilson, N. A. Ali, S. Muralidharan, M. Mirzaei, M. Mariani, G. Assadourian, A. Lee, S. C. van Sluyter, and P. A. Haynes, "Less label, more free: Approaches in label-free quantitative mass spectrometry," *PROTEOMICS*, vol. 11, pp. 535–553, Feb. 2011.

[100] Y. Zhang, Z. Wen, M. P. Washburn, and L. Florens, "Improving Label-Free Quantitative Proteomics Strategies by Distributing Shared Peptides and Stabilizing Variance," *Analytical Chemistry*, vol. 87, pp. 4749–4756, May 2015.

[101] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann, "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ," *Molecular & cellular proteomics: MCP*, vol. 13, pp. 2513–2526, Sept. 2014.

[102] B. Schilling, M. J. Rardin, B. X. MacLean, A. M. Zawadzka, B. E. Frewen, M. P. Cusack, D. J. Sorensen, M. S. Bereman, E. Jing, C. C. Wu, E. Verdin, C. R. Kahn, M. J. Maccoss, and B. W. Gibson, "Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: Application to protein acetylation and phosphorylation," *Molecular &*

cellular proteomics: MCP, vol. 11, pp. 202–214, May 2012.

[103] C. R. Kinsinger, J. Apffel, M. Baker, X. Bian, R. Bradshaw, M.-Y. Brusniak, D. W. Chan, E. W. Deutsch, J. Gorman, R. Grimm, W. Hancock, H. Hermjakob, D. Horn, C. Hunter, P. Kolar, H.-J. Kraus, H. Langen, R. Linding, L. Moritz, G. S. Omenn, R. Orlando, A. Pandey, P. Ping, A. Rahbar, R. Rivers, S. L. Seymour, R. J. Simpson, D. Slotta, R. D. Smith, S. E. Stein, D. L. Tabb, D. Tagle, J. R. Y. Iii, and H. Rodriguez, "Recommendations for Mass Spectrometry Data Quality Metrics for Open Access Data (Corollary to the Amsterdam Principles)," p. 14, 2013.

[104] J. K. Eng, A. L. McCormack, and J. R. Yates, "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database," *Journal of the American Society for Mass Spectrometry*, vol. 5, pp. 976–989, Nov. 1994.

[105] D. N. Perkins, D. J. Pappin, D. M. Creasy, and J. S. Cottrell, "Probability-based protein identification by searching sequence databases using mass spectrometry data," *Electrophoresis*, vol. 20, pp. 3551–3567, Dec. 1999.

[106] M. J. MacCoss, C. C. Wu, and J. R. Yates, "Probability-Based Validation of Protein Identifications Using a Modified SEQUEST Algorithm," *Analytical Chemistry*, vol. 74, pp. 5593–5599, Nov. 2002.

[107] A. Keller, A. I. Nesvizhskii, E. Kolker, and R. Aebersold, "Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search," *Analytical Chemistry*, vol. 74, pp. 5383–5392, Oct. 2002.

[108] L. Käll, J. D. Canterbury, J. Weston, W. S. Noble, and M. J. MacCoss, "Semi-supervised learning for peptide identification from shotgun proteomics datasets," *Nature Methods*, vol. 4, pp. 923–925, Nov. 2007.

[109] T. Taus, T. Köcher, P. Pichler, C. Paschke, A. Schmidt, C. Henrich, and K. Mechtler, "Universal and Confident Phosphorylation Site Localization Using phosphoRS," *Journal of Proteome Research*, vol. 10, pp. 5354–5362, Dec. 2011.

[110] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature*

Protocols, vol. 4, no. 1, pp. 44–57, 2009.

[111] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[112] J. C. van der Mijn, M. Labots, S. R. Piersma, T. V. Pham, J. C. Knol, H. J. Broxterman, H. M. Verheul, and C. R. Jiménez, "Evaluation of different phosphotyrosine antibodies for label-free phosphoproteomics," *Journal of Proteomics*, vol. 127, pp. 259–263, Sept. 2015.

[113] Y. Abe, M. Nagano, A. Tada, J. Adachi, and T. Tomonaga, "Deep Phosphotyrosine Proteomics by Optimization of Phosphotyrosine Enrichment and MS/MS Parameters," *Journal of Proteome Research*, vol. 16, pp. 1077–1086, Feb. 2017.

[114] A. N. Kettenbach and S. A. Gerber, "Rapid and Reproducible Single-Stage Phosphopeptide Enrichment of Complex Peptide Mixtures: Application to General and Phosphotyrosine-Specific Phosphoproteomics Experiments," *Analytical Chemistry*, vol. 83, pp. 7635–7644, Oct. 2011.

[115] V. Chumbalkar, K. Latha, Y. Hwang, R. Maywald, L. Hawley, R. Sawaya, L. Diao, K. Baggerly, W. K. Cavenee, F. B. Furnari, and O. Bogler, "Analysis of Phosphotyrosine Signaling in Glioblastoma Identifies STAT5 as a Novel Downstream Target of ΔEGFR," *Journal of Proteome Research*, vol. 10, pp. 1343–1352, Mar. 2011.

[116] M. Tinti, A. P. Nardozza, E. Ferrari, F. Sacco, S. Corallino, L. Castagnoli, and G. Cesareni, "The 4G10, pY20 and p-TYR-100 antibody specificity: Profiling by peptide microarrays," *New Biotechnology*, vol. 29, pp. 571–577, June 2012.

[117] G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner, "WebLogo: A sequence logo generator," *Genome Research*, vol. 14, pp. 1188–1190, June 2004.

[118] R. Amanchy, K. Kandasamy, S. Mathivanan, B. Periaswamy, R. Reddy, W.-H. Yoon, J. Joore, M. A. Beer, L. Cope, and A. Pandey, "Identification of Novel Phosphorylation Motifs Through an Integrative Computational and Experimental Analysis of the Human Phosphoproteome," *Journal of proteomics & bioinformatics*, vol. 4, no. 2, pp. 22–35, 2011.

[119] Y. Zhang, A. Wolf-Yadlin, P. L. Ross, D. J. Pappin, J. Rush, D. A. Lauffenburger, and F. M. White, "Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules," *Molecular & Cellular Proteomics*, vol. 4, pp. 1240–1250, Jan. 2005.

[120] M. Cargnello and P. P. Roux, "Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases," *Microbiology and Molecular Biology Reviews : MMBR*, vol. 75, pp. 50–83, Mar. 2011.

[121] N. T. Snider and M. B. Omary, "Post-translational modifications of intermediate filament proteins: Mechanisms and functions," *Nature reviews. Molecular cell biology*, vol. 15, pp. 163–177, Mar. 2014.

[122] J. Y. Baik, M. S. Lee, S. R. An, S. K. Yoon, E. J. Joo, Y. H. Kim, H. W. Park, and G. M. Lee, "Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin," *Biotechnology and Bioengineering*, vol. 93, pp. 361–371, Feb. 2006.

[123] L. R. Forbes, J. Milner, and E. Haddad, "STAT3: A year in Review," *Current opinion in hematology*, vol. 23, pp. 23–27, Jan. 2016.