Profiling the Dynamic Changes in the Phosphoproteome of Chinese Hamster Ovary Cells

PhD

Prashant – MTech

January 2020

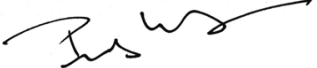The research described in this thesis was conducted under the supervision of

Dr Paula Meleady

National Institute for Cellular Biotechnology,

School of Biotechnology,

Dublin City University

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of PhD is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ (Candidate)   ID No.: 15212429

Date: 26th August 2020

This thesis is dedicated to my late grandmother Chaman Devi.

Acknowledgments

Overview of publications presented in this thesis:

Publication title	Journal	Authors	Contribution
The Expression Pattern of the Phosphoproteome Is Significantly Changed During the Growth Phases of Recombinant CHO Cell Culture	Biotechnology journal doi: 04 August 2018 Volume13, Issue10 *DOI:10.1002/biot.201700221*	**Prashant Kaushik,** Henry M, Clynes M, Meleady P.	First co-author, investigation and manuscript preparation.
Increased mAb production in amplified CHO cell lines is associated with increased interaction of CREB1 with transgene promoter	Current Research in Biotechnology 05 October 2019 Volume 1, November 2019, Pages 49-5 *DOI:10.1016/j.crbiot.2019.09.001*	Hussain Dahodwala, **Prashant Kaushik** Tejwani V, Chih-ChungKuo, Menard P, Henry M, Voldborg BG, Lewis NE, Meleady P, Sharfstein ST,	Joint first co-author, investigation and manuscript preparation.
Quantitative label-free LC-MS/MS comparative proteomic and phosphoproteomic analysis of CHO-K1 cells adapted to growth in glutamine-free media.	Biotechnology Letters 29 June 2020 *DOI:10.1007/s10529-020-02953-7*	**Prashant Kaushik**, Ricardo Valdés-Bango Curell, Michael Henry, Niall Barron, Paula Meleady	First co-author, investigation and manuscript preparation.

The role of this candidate stated here is accurate:

Primary supervisor   Signed: *Paula Meleady*   Date: 27th August 2020

Student   Signed: *[signature]*   Date: 26/August/2020

Publications not included in this thesis:

1. **Phosphopeptide enrichment and LC-MS/MS analysis to study the phosphoproteome of recombinant Chinese Hamster Ovary cells**. Henry M, Coleman O, **Prashant**, Clynes M, Meleady P. Methods in Molecular Biology (2017). Volume 1603, pp 195-208.

2. **Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift.** Henry M, Power M, **Kaushik P,** Coleman O, Clynes M, Meleady P. Journal of Proteome Research (2017). Volume 16(7) pp 2339-2358.

3. **Proteogenomic Annotation of Chinese Hamsters Reveals Extensive Novel Translation Events and Endogenous Retroviral Elements.** Li S, Cha SW, Heffner K, Hizal DB, Bowen MA, Chaerkady R, Cole RN, Tejwani V, **Kaushik P**, Henry M, Meleady P, Sharfstein ST, Betenbaugh MJ, Bafna V, Lewis NE. Journal of Proteome Research (2019). Volume 18(6) pp 2433-2445

Achievements

Invited to review manuscript for Biotechnology Letters: 1st Oct 2019

ESACT UK travel grant for oral presentation: 9th Jan 2019

Biological Research Society best oral presentation: 25th Jan 2019

Irish Mass Spectrometry Society best poster prize: 16th May 2019

ESACT Lausanne travel grant awarded for poster presentation: 14th May 2017

Table of Contents

Abbreviations

2DIGE	Two-dimensional difference gel electrophoresis
AC	Alternating Current
ACN	Acetonitrile
AUC	Area under the curve
C-trap	Curved linear trap for ion injection
CHAPS	3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate
CHEA	α-Cyano-4-hydroxycinnamic acid
CHO	Chinese Hamster Ovary
CID	Collison Induced Dissociation
CREB1	Cyclic response element binding protein
DAVID	Database for annotation, visualization and integrated discovery
DDA	Data dependent acquisition
DTT	Dithiothreitol
ECD	Electron capture dissociation
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
ETD	Electron transfer dissociation
EThcD	Electron-transfer/higher-energy collision dissociation
FTMS	Fourier transform Ion cyclotron resonance mass Spectrometer
GO	Gene ontology
HCD	Higher energy collisional dissociation
ICAT	Isotope-coded affinity tag
IDA	Iminodiacetate
IgG	Immunoglobulin G
IMAC	Immobilized metal affinity chromatography
IPA	Ingenuity pathway analysis
iTRAQ	Isobaric tags for relative and absolute quantitation
KEGG	Kyoto encyclopaedia of genes and genomes
LC	Liquid chromatography

LC-MS	Liquid chromatography coupled with mass spectrometry
LTQ	Linear ion trap mass spectrometer
LysC	Lysine-sensitive aspartokinase 3
mAb	Monoclonal Antibody
MALDI	Matrix assisted laser desorption ionization
MOAC	Metal oxide affinity chromatography
mRNA	Messenger ribonucleic acid
MS	Mass spectrometer
MS/MS	Sequential mass detection
MSA	Multistage activation
nano-ESI	nano-Electrospray ionization
NTA	Nitrilotriacetate
*p	Phospho
PTMs	Posttranslational modifications
rCHO	Recombinant Chinese hamster ovary
RefSeq	Reference Sequence
REP	Replicate
RF	Radio frequency
Ribo-Seq	Ribosome sequencing
SDS	Sodium dodecyl sulphate
Ser	Serine
SH	SRC Homology
SILAC	Stable isotope labelling by/with amino acids in cell culture
STAT	Signal transducer and transcription activator
Thr	Threonine
TiO2	Titanium dioxide
TOF	Time of flight
Tyr	Tyrosine
UHPLC	Ultra-high performance liquid chromatography
z-Score	Standard score

Abstract

Title: Profiling the Dynamic Changes in the Phosphoproteome of Chinese Hamster Ovary Cells

Author: Prashant

Post-translational modification of proteins by reversible phosphorylation plays a pivotal role in regulating vital cellular processes. Despite the importance of the phosphorylation level of regulation, little work has been carried out on the phosphoproteomic characterization of Chinese hamster ovary (CHO) cells in bioprocess-relevant conditions. To bridge this knowledge gap, through application of mass spectrometry-based proteomics and phosphoproteomics, I present three original, studies that investigate (1) the dynamic nature of the CHO cell phosphoproteome in response to changing culture conditions; (2) differential activation of transcription factors between cell lines of varying specific productivity; and (3) changes in the CHO cell proteome and phosphoproteome in response to adaptation to growth in glutamine free media.

Study 1 shows that the inclusion of phosphoproteomic data significantly improves proteome coverage and gives insights into cell signalling pathways that could be targeted to control growth. Study 2 shows that the nuclear proteome and phosphoproteome have an essential role in regulating the final productivity of recombinant proteins from CHO cells and that CREB1 may play a role in transcriptional enhancement. Finally, study 3 provides a comprehensive proteomic and phosphoproteomic analysis of how cellular proteome expression changes after adaptation to glutamine-free growth conditions and highlights critical pathways to consider when designing future studies to further understand and engineer glutamine metabolism, and rational design of improved feeding strategies.

Together, these studies advance our understanding of CHO cell biology and provide new avenues for exploration of targets for cell line engineering to improve the efficiency of production of recombinant biotherapeutics

1. Chapter 1: Application of proteomics throughout this thesis and an introduction to Chinese hamster ovary cells.

1.1 Introduction to proteomics

Proteomics is essentially the scientific investigation concerning protein or proteins in a biological system at a given time. The term proteome was first used in 1995 to describe the protein complement to the genome (Wilkins et al., 1996). Since then, the limitations and the scopes on how proteins are studied have changed, which in turn has defined the term proteomics as a system-wide identification and analysis of protein sequence, localization, abundance, post-translational modifications and biomolecular interactions (Käll & Vitek, 2011). In the quest to answer a biological question through proteomics, it must be kept in mind that the proteome of a cell gives an immediate account of the environment in which it is being studied. Hence, considering all the possibilities that can influence a cellular biological environment, it is likely that any given genome could give rise to infinite proteomes. Proteins are initially synthesized as polypeptide chains of amino acids by the cellular translational machinery and later folded into complex quaternary structures. The translation machinery utilizes the amino acid sequence information encoded in an mRNA molecule, which acts as a medium to translate data from the protein-coding DNA sequence to a polypeptide product. Information about all the proteins in a biological system is encoded in the genome as protein coding genes; sequencing the genome however, does not give a complete account of the cellular proteomic state. This is primarily because 1), not all protein-coding DNA sequences are translated into an mRNA, 2) not all mRNA molecules are translated into protein sequence at the same rate, and 3) one gene can give rise to several mRNAs, and 4) one mRNA can also produce multiple variants of

a protein. Therefore, studying only the genomic and transcriptome information is likely to give an incomplete understanding of a biological system as a whole. Studying proteins provides a potentially more realistic account of the conditions under which a biological system is being investigated.

Proteins are elegant and sophisticated molecules that perform highly specialized functions. The ability of proteins to act as final effector molecules is dependent ultimately on the physicochemical properties that allow them to partake in their specialized role. As an example, the SH3 domain on a protein is a motif of 50 amino acid residues that has an affinity towards binding to proteins that contain specific proline-rich sequences (Kurochkina & Guha, 2013), thus facilitating protein-protein interaction of only particular targets; whereas a closely related SH2 domain binds to proteins with phosphorylated tyrosine residues (Liu, 2017). Information like this is utilized by computational algorithms to generate protein-protein interaction networks, thus providing networks or pathways of proteins that contain a chain of events from a stimulus to an effect. To utilize such computational platforms, we first must know which proteins are being expressed in cells under the specific conditions or stimulus (qualitative proteomics) and does the state or the stimulus have any effect on the expression levels of these proteins (quantitative proteomics). These two information levels can give us a great insight into the mechanisms at how the cellular machinery utilizes a mostly static genome to achieve a high level of biological flexibility through differential expression of proteins. Specialized domains such as SH3 and SH2, are encoded in the genome and are predetermined before protein synthesis, i.e., translation. Cellular proteins also undergo a variety of modifications after protein synthesis (post-translational modifications); these modifications can have a profound impact on the activity of a protein and affect almost all cellular processes, including bioprocess

relevant processes such as CHO cell growth, apoptosis, and specific productivity. This thesis explores the landscape of post-translational modification of cellular proteins in CHO cells in appropriate bioprocess contexts, explicitly concentrating on phosphorylation, and through the discussion of these findings, makes a case on the importance of studying alterations to post-translational modifications of cellular proteins in CHO cells.

1.2   Protein post-translational modification

Properties that provide functional relevance to a protein such as the activity state, localization, turnover rate, solubility and interactions with other proteins do not only depend on its amino acid sequence and structure but also how these amino acids are post-translationally modified. Post-translational modification, either by the proteolytic cleavage or by the covalent modification of amino acids, brings about substantial changes in the functionality of proteins. Where changes in the genome happen at an evolutionary scale, PTMs on proteins allow cells to respond to the internal and the external stimulus 'on the fly' (Prabakaran, Lippens, Steen, & Gunawardena, 2012). Modification by phosphorylation, for example, can turn signaling cascades on or off. After synthesis, a protein can be modified in several ways. Some of the bioprocess-relevant post-translational modifications include glycosylation, ubiquitination and phosphorylation. (1) Glycosylation is a covalent attachment of a sugar residue to a protein through enzymatic action. It is highly relevant in bioprocessing, firstly because glycoproteins are an integral part of the cellular membrane, cytoplasm, and the nucleolus, providing functional and structural activity (Spiro, 2002). Secondly, because correct glycosylation of the recombinant protein profoundly impacts the biological activity, function, clearance from circulation, and antigenicity of a therapeutic protein (Tayi & Butler, 2015); hence, the choice of the host system for

production of recombinant therapeutic product largely depends on its ability to mimic the human-like glycosylation profile. (2) Ubiquitination, the covalent attachment of ubiquitin, a highly conserved 76 amino acid polypeptides to a variety of cellular proteins, is a significant pathway for selective proteolysis of proteins, thus maintaining cellular homeostasis (Haas & Siepmann, 1997). Ectopic ubiquitination signalling can lead to the accumulation and aggregation of intracellular proteins which has been linked to variety of pathological outcomes such as cancer, metabolic syndromes, neurodegenerative diseases, autoimmunity, inflammatory disorders, infection and muscle dystrophies (Popovic, Vucic, & Dikic, 2014); therefore the ubiquitination system is an emerging target for cancer therapies (Veggiani, Gerpe, Sidhu, & Zhang, 2019). (3) Phosphorylation, the reversible phosphorylation of serine, threonine and tyrosine amino acids, is a crucial regulatory post-translational modification of proteins. The regulatory switch provided by phosphorylation of proteins virtually affects every cellular process, including growth, cell division, differentiation, membrane transport, etc. (Ubersax & Ferrell, 2007). The work described in this thesis specifically explores the dynamics of the phosphorylation level of regulation in CHO cells. Chapter 2 in this thesis shows the temporal changes in the phosphoproteome of IgG-producing CHO DP12 cells in a 4-day suspension culture. Chapter 3 focuses specifically on the nuclear phosphoproteome of IgG-producing CHO DG44 cells and integrates this information with transcriptomic and nuclear proteomic data to understand the activation of transcription factors in high productivity cell lines. Chapter 4 provides an insight into glutamine metabolism in CHO cells by comparing the proteomic and phosphoproteomic landscape of CHO-K1 cells adapted to growth in glutamine free media to a parental CHO-K1 cell line.

1.3 Cell signalling regulation by protein phosphorylation.

Cell signalling is a biological process by which cells respond to changes to their internal and external environment. For example, the triggering of gene expression by binding of an extracellular ligand to a cell surface receptor or cellular migration motivated by osmotic flux is an example of how cell signalling regulates complex biochemical and biomechanical processes, and reversible phosphorylation of proteins is a major switch mechanism for cell signalling control.

Phosphorylation of proteins by kinases converts the polar neutral side chain into a negatively charged tetrahedral phosphate. The addition of a small negative charge profoundly affects a protein's microenvironment by perturbing the energy landscape of a protein and drives conformational changes within protein secondary structures (Groban, Narayanan, & Jacobson, 2006). These conformational changes create, expose or mask active sites for protein-protein interactions (Stultz, Levin, & Edelman, 2002). On the other hand, enzyme phosphatases remove the phosphate groups, thus reversing the activity of kinases. This synchronization between the kinases, the phosphoproteins, and the phosphatases at a systemic level is fundamental to regulation by cell signalling.

The importance of system-wide regulation by phosphorylation can be estimated by the fact that up to 2% of the eukaryote genome accounts for genes that codes for proteins of the kinase gene family (Y. Zhang, Fonslow, Shan, Baek, & Yates, 2013). Furthermore, up to 30% of all cellular proteins are estimated to be phosphorylated on at least one residue at any given time (Philip Cohen, 2000). Since the approval of the first therapeutic kinase inhibitor Imatinib for the treatment of myeloid leukaemia in 2001, 37 drugs that target protein kinases have received FDA approval, and about 150 other kinase targeting drugs are in clinical trials (Bhullar et al., 2018). Despite the

widespread phosphorylation of proteins, a functionally phosphorylated isoform of a regulatory protein is expressed at significantly lower levels than the total amount of the protein present in the sample. As a result, the large-scale analysis of phosphoproteins is achieved by phosphoprotein/phosphopeptide enrichment strategies (Goshe, 2006) described in section 1.4 of this chapter followed by the mass spectrometry analysis of these "phospho-enriched" sample fractions described in detail in section 1.4.3.

In brief, the fundamentals of mass spectrometry is the measurement of m/z (mass to charge ratio) of a gas phase peptide ion. Achieving large scale quantitative proteomics requires meticulous sample preparation, in-depth understanding of the MS instrumentation and complementing computational platforms. The following sections provide an overview of a typical proteomic workflow, and the proteomic methodology explicitly applied to this thesis have contributed to the publication of a methodology book chapter entitled "Phosphopeptide Enrichment and LC-MS/MS Analysis to Study the Phosphoproteome of Recombinant Chinese Hamster Ovary Cells" (Henry et al., 2017a).

1.4   Mass spectrometry-based proteomics

The cellular proteome is a complex mixture with an estimated 2-4 million protein molecules per cubic micron (Milo, 2013). Given that the diameter of recombinant CHO (rCHO) cell in suspension culture is estimated to be ~ 15µm (Y. Han et al., 2006) and an assumed perfect spherical shape, a single CHO cell could potentially have 2.8×10^9 to 5.7×10^9 proteins. Furthermore, the dynamic range of proteins in a cell ranges from one copy per cell to ten million copies per cell (Zubarev, 2013) . Therefore, early proteomic techniques relied on the isolation of a single protein or a group of proteins from a complex mixture of the cellular proteome. Although these techniques,

6

such as gel-based protein separation and western blotting, have proved to be extremely useful in identifying critical cellular proteins and are still in widespread use, they are severely limited in the ability to generate signalling pathways to complement the current computational resources available for system-wide analysis of cellular proteomics. Mass spectrometry has become the method of choice for analysis of complex protein samples. Recent developments in mass spectrometry instrumentation and supporting computation have enabled detection of thousands of proteins and their modifications such as phosphorylation with high accuracy and precision. This thesis utilizes state-of-the-art mass spectrometry to identify differential expression of protein and differential site-specific phosphorylation on CHO cells.

1.4.1 Proteomic sample preparation for Mass Spectrometry.

Due to the dynamic nature of the proteome and labile post-translational modifications, sample preparation is essentially the most crucial step in any proteomics study on which the downstream analysis relies on. One advantage of working with suspension CHO cell cultures is having some degree of control over the amount of starting material or cell number (i.e. cell numbers can be scaled up for proteomic sample preparation). In my experience, 1×10^7 CHO cells generates approximately 1 mg protein of lysate, which is a sufficient amount of starting material for phosphopeptide enrichment and parallel total protein analysis. Low starting material is often a challenge with limited amounts of clinical sample material. Regardless of the sample origin, a proteomic workflow typically requires lysis of cells or tissue to release the intracellular proteins into a solution. For CHO cell suspension cultures, cell lysis can be achieved by either mechanical lysis, i.e., by using sonication, or chemical lysis by using a specialized lysing agent such as SDS surfactant. Depending on the downstream application, attention needs to be given to the chemical composition of the buffer in

which proteins are solubilized. It is critical to have a near intracellular pH to prevent pH-dependent denaturation of proteins during the lysis process, and also to achieve high solubility of proteins in the lysis buffer. A high concentration of potent chaotropic agents such as urea or thiourea is often used to disrupt hydrogen bonding networks between water molecules and protein residues, thus reducing the stability of the native state of proteins by weakening the hydrophobic interactions; this facilitates the solubilisation of otherwise insoluble proteins (Salvi et al., 2005). Detergents are often used to disrupt the hydrophobic interactions between and within proteins. The utilisation of zwitterionic detergents such as CHAPS provides higher solubility for hydrophobic membrane proteins (Luche et al., 2003). Sodium Dodecyl Sulphate (SDS), an anionic detergent, is often a detergent of choice where rapid solubilization of proteins is required (Wiśniewski et al., 2009). The selection of the appropriate lysis buffer is a crucial step; what suits one application might not be suitable for another. For example, detergents such as SDS provide excellent solubility of proteins and are compatible with 2D gel electrophoresis but are a significant nuisance for high-resolution mass spectrometry. Detergent contamination of the mass spectrometer is a costly affair, often requiring replacement of columns/tubing, etc. (Yeung et al., 2008). Mass spectrometry compatible detergents exist (e.g. Rapigest from Waters and ProteaseMax from Promega), but their high cost limits their regular usage. To avoid detergent associated complexities, and to minimize the sample handling steps to secure the labile phosphorylation modification, the work in this thesis was carried out by mechanical lysis using a sonicator. To prevent post lysis removal of phosphorylation modifications by release of intracellular phosphatases, a commercial chemical cocktail of phosphatase inhibitors (Halt™ Protease Inhibitor, Thermo Fischer Scientific) was added at the lysis stage of the process.

The preparation of a protein sample is a multi-step process. Depending on the complexity of the sample, analysis by MS could be "top-down" or "bottom-up." Top-down proteomics analyses intact proteins for full characterization. However, the application is limited to users in the study of a single protein or simple protein mixtures (Resemann et al., 2010). Top-down proteomics is of particular importance in bioproduction for the characterization of the critical quality attributes of a recombinant protein product (Mariñoet al., 2010). For complex protein samples such as whole cell lysates or tissue lysates, proteins are first denatured and are then enzymatically digested by proteolytic enzymes to convert complex protein mixture into easily ionizable peptides. However, complete digestion of proteins is prevented by the steric hindrance provided by complex protein quaternary structures; hence for bottom-up proteomics, cell or tissue protein samples are obtained in denaturing buffers containing chaotropic agents such as urea or thiourea which promotes the disruption of the disulfide bonds of a quaternary structure to yield single-stranded polypeptides. Furthermore, denaturation of the protein sample is facilitated by treatment with potent reducing agents such as dithiothreitol (DTT). After denaturation by DTT, exposed sulfhydryl groups are then alkylated with iodoacetamide to prevent disulfide bonds reforming (Zhang et al., 2013). The proteolytic enzymes, which cleave at specific amino acid residues, are then used to "chop-down" large polypeptides to smaller peptide fragments. Trypsin is the most widely used protease for protein digestion, although other proteases such as chymotrypsin, LysC, LysN, AspN, GluC and ArgC are also available (Giansanti, Tsiatsiani, Low, & Heck, 2016). Trypsin cleaves exclusively at the C-terminal of lysine and arginine and allows the formation of peptides ranging from 700 to 1500 Daltons, an ideal size for MS analysis. However, the presence of a proline residue or a phosphorylated amino acid close to the cleavage

site can reduce trypsin activity; therefore, digestion efficiency is increased either by increasing the amount of trypsin used or by using multiple proteases with different specificities such as a combination of trypsin and LysC. The work presented in this thesis uses a trypsin double digestion approach in which samples are treated with 1:100 (enzyme:substrate) trypsin twice in the 12-hour incubation period. As shown in Figure 1.1, the number of highly charged peptides were significantly reduced by using a trypsin double digestion approach as an indication of efficient digestion of the cellular proteome (Sun at al., 2004).
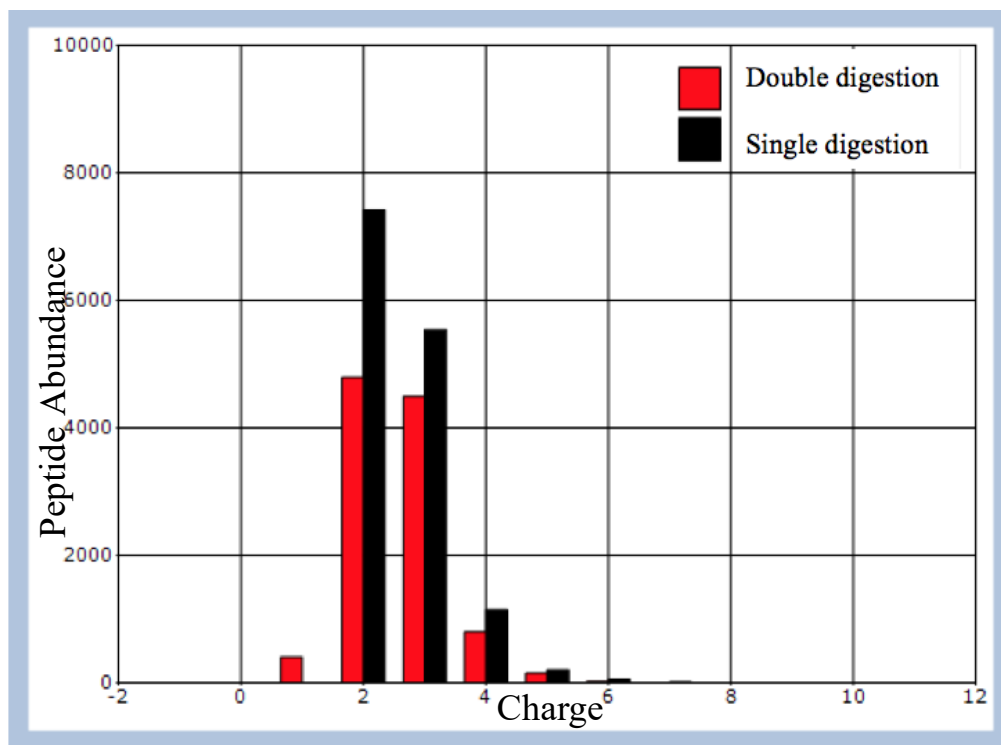
Figure 1.1 The graph shows the distribution of the charge state per phosphopeptide identified. Sample in black represents phosphopeptide enrichment from CHO DP12 whole cell lysate using single trypsin digestion (1:100). Sample in red represents phosphopeptides enrichment after trypsin double-digestion. More efficient digestion is achieved by using double trypsin digestion by reducing the number of multiply charged phosphopeptides. Both samples were enriched using Fe_NTA chromatography.

1.4.2 Subcellular proteome analysis.

To increase proteome coverage by mass spectrometry, the whole cell proteome is often divided into subcellular fractions. Division of a complex whole-cell proteome into its constituent organelle proteome also provides a more comprehensive understanding of residing resident proteins. The exploration of the proteomes regarding cellular compartments is a practical strategy for understanding subcellular proteome associated specialised cellular functions such as protein translocation and activation. Biochemical techniques for subcellular fractionation coupled with phosphopeptide enrichment for individual fractions will possibly bring about the detection of trace phosphoproteins that may not show up from phosphopeptide enrichment of a whole-cell lysate. Furthermore, a hallmark of many signalling pathways is the spatial and temporal regulation of both protein kinases, phosphatases and phosphoproteins to ensure an appropriate balance of protein phosphorylation (Bauman & Scott, 2002). Hence, it becomes essential to study phosphorylation in context with cellular compartments. Spatial distribution of kinases and phosphatases also provide the gradient for substrate phosphorylation (Kholodenko, 2002). Phosphorylation also facilitates the translocation of phosphoproteins from one cellular compartment to another. For example, many transcription activators such as STAT migrate to the nucleus after phosphorylation by an upstream kinase (Jak-1 for STAT) in the cytoplasm (Decker & Kovarik, 2000), therefore studying the phosphorylation stoichiometry of STAT from a whole cell lysate will provide an unreliable account of STAT activation.

For subcellular phosphoproteomics, the organelle for whose proteome is to be studied is enriched from the cell. The most common approach to achieve a subcellular proteome is by fractionation. Typically, subcellular fractionation is performed by

differential density centrifugation. Cells are disrupted to keep the internal organelles intact, mainly by the use of detergents or hypo-osmotic shock. The organelles are then separated under continuous or discontinuous gradients using viscous media, such as sucrose of different osmolality, viscosities or densities. Due to the difference in size and density, different organelles sediment at a density gradient that matches their internal density. Due to the complex nature of fractionation through centrifugation, a number of commercial kits are available to fractionate cellular components without elaborate centrifugation steps. Investigating the proteome of a cell organelle allows a better understanding of both proteins and the organelles in question. Sub-cellular enrichment techniques are mostly biochemical, and due to high chances of contamination, the purity of the enriched compartment is checked before downstream application (Walther & Mann, 2010). For example, enrichment of a nuclear fraction can be verified by western blotting of the enriched and non-enriched fraction using antibodies against marker proteins from an enriched cellular component. As an example, Figure 1.2 shows the enrichment of the nuclear fraction of CHO cells using an Active Motif Nuclear Extract Kit that uses a high salt extraction method by suspending cells in a hypotonic buffer which causes cell swelling and weakening of the plasma membrane. Detergents in the buffer cause cytoplasmic proteins to solubilize and leak into the supernatant. After removal of the supernatant, nuclei are lysed separately. Enrichment of the nuclear fraction was validated by confirming the presence of a nuclear marker protein, e.g. Lamin A, and the absence of a cytosolic marker protein, e.g. HSP 60, in the nuclear-enriched fraction by western blot analysis.

12
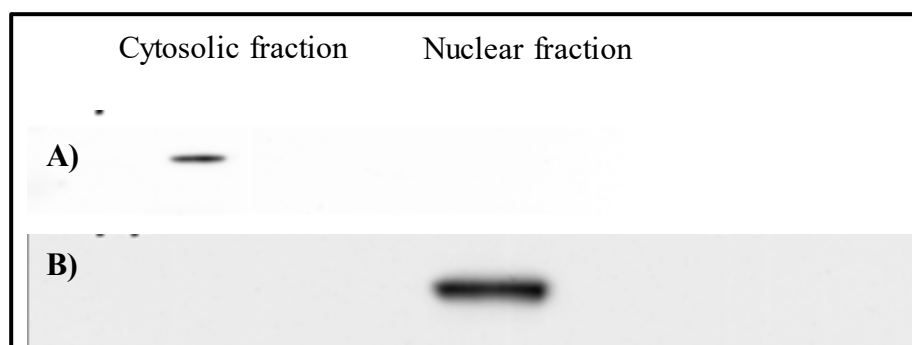
Figure 1.2 **Western Blot analysis of nuclear protein fraction enrichment.** *The nuclear proteom... ...alidation of enric... ...rker.*

Chapte... ...roteome and ph... ...cription factors. Nuclear proteomic and phosphoproteomic information was then integrated with transcriptomic data to elucidate the activation of specific transcription factors in cell lines with high specific productivity (Dahodwala et al., 2019a).

1.4.3   Phosphopeptide enrichment methods

To deal with the sub-stoichiometric nature of phosphorylated proteins, phosphoproteomic analysis requires the selective enrichment of phosphopeptides from the whole-cell proteome (Paulo et at., 2018). Several techniques are available to enrich phosphopeptides from the complex mixture of a whole-cell proteome digest. Broadly, the three primary methods that are currently used for phosphopeptide enrichment are IMAC (Immobilized Metal affinity chromatography), MOAC (Metal oxide affinity chromatography) and enrichment of phosphopeptides with immune-affinity based techniques (Beltran & Cutillas, 2012).

Alternative enrichment method of chemical derivatization by the beta-elimination of the phosphate group of serine and threonine residues followed by Michael addition reaction to link a phosphosite firmly to a solid support is also available (Knight et al., 2003). However, due to the high non-specific reactivity, non-compatibility with p-Tyr

and high complexity, phosphopeptide enrichment by chemical derivatization is not a popular technique (Nika et al., 2013).

1.4.4 IMAC

Initially introduced for the enrichment of phosphoproteins in 1986 (Andersson & Porath, 1986) IMAC is now the most the widely used technique for phosphopeptide enrichment (Thingholm & Larsen, 2016a). Phosphopeptide enrichment by IMAC is based on the affinity between the negatively charged phosphate group and the positively charged metal ions such as Fe^{3+} or Ga^{3+} that are chelated to a solid support such as beads coated with Nitrilotriacetate (NTA) or Iminodiacetate (IDA) which act as the IMAC resins. For phosphopeptide enrichment by IMAC, Fe^{3+}_NTA, ion-resin complex is the most widely used combination (Ye et al., 2010a).

There are a number of challenges associated with phosphopeptide enrichment by IMAC. The specificity of IMAC enrichment depends upon many factors such as (1) the purity of ions used for matrix preparation (Ye et al., 2010b), (2) the presence of contaminating agents such as nucleic acids which potentially binds to the IMAC matrix resulting in a loss of phosphopeptides (Yaojun Li et al., 2009), and (3) selection of proteases. The gold standard for shotgun proteomics involves the use of the protease trypsin, which cleaves at the N-terminal of lysine and arginine. Tryptic peptides generated by trypsin may contain multiple acidic residues (Evnin et al., 1990) whereas the protease Glu-C cleaves at C-terminal of glutamic acid and aspartic acid residues, giving rise to peptides with single acidic residues only. As non-specific binding of acidic peptides is an issue in IMAC, the application of Glu-C as a protease has been shown to improve enrichment specificity (Riggs et al., 2005).

To reduce the non-specific binding of the acidic residues to the metal ion, the pH of the loading buffer, i.e. pH 2-2.5, is carefully selected based on the difference between

the acidic dissociation constant (pKa) between acidic residues of glutamic acid (pKa, 3.65) and aspartic acid (pKa, 4.25) and the phosphate group (pKa, pH2.1). Protonation of acidic residues mask the negative charge of the carboxyl group and do not bind to the positively charged metal ion. On the other hand, the phosphate group would still be deprotonated and be carrying a negative charge which could attach to the positive metal ion (Fílla & Honys, 2012).

The disadvantages of IMAC phosphopeptide enrichment are its low specificity in comparison to other metal affinities based techniques such as MOAC (Kweon & Håkansson, 2006) and its low tolerance for detergents and EDTA (Aryal & Ross, 2010). Also, due to the non-covalent interaction between metal ion and substrate, there is a possibility of contamination of phosphopeptides by leaching metal ions (Cheung, Wong, & Ng, 2012).

Despite the challenges and disadvantages mentioned above, phosphopeptide enrichment by IMAC is still a widely used technique and is typically used in combination with other separation techniques such as pre-fractionation of whole-cell peptides by strong cation exchange chromatography (Beltran & Cutillas, 2012).

1.4.5   Metal Oxide Affinity chromatography

Phosphopeptide enrichment by MOAC utilizes the affinity of positively charged metal oxide toward the negatively charged phosphate group. The significant advantage of MOAC over IMAC is the high stability of metal oxide than traditional silica-based stationary phase (Nawrocki et al., 2004). A variety of metal oxides are available for phosphopeptide enrichment such as titanium dioxide (TiO_2), zirconium dioxide (ZrO_2), aluminium oxide (Al_2O_3), and niobium oxide (Nb_2O_5). Currently, TiO_2 one of the most favoured phosphopeptide enrichment technique (Thingholm & Larsen, 2016c).

At acidic pH, TiO_2 exhibit positive charges on the surface, which facilitates the absorption of the negatively charged phosphate group, and phosphopeptide enrichment from whole-cell lysates using TiO_2 has been reported to bind the phosphopeptides over acidic peptides preferentially, showing high selectivity and reproducibility (Larsen et al., 2005). For phosphopeptide enrichment, the utility of TiO_2 has been tested in an on-line and an off-line fashion to address the widely reported issue of low reproducibility of phosphopeptide enrichment. In one study phosphopeptides were eluted using an LC auto-sampler on a first-dimension column containing TiO_2 resin packed with fused silica onto a second trap column, from which phosphopeptides were eluted using a reverse-phase gradient thus successfully enriching phosphopeptides from 1pmol of starting material (Cantin et al., 2007). In a more recent study, magnetic hyper-porous TiO_2 polymer matrix was coupled to a magnetic particle processing robot and achieved high-throughput, reproducible, high-fidelity phosphopeptide enrichment (>90% phosphopeptide purity) (Tape et al., 2014). In a milestone study, more than 50,000 distinct phosphorylated peptides were identified in a hyper-stimulated human cancer cell line by pre-fractionation of the whole cell lysate by strong cation exchange chromatography before enrichment by TiO_2 chromatography and using a quadrupole Orbitrap instrument (Q Exactive) mass spectrometer (Sharma et al., 2014). New state-of-the-art instruments such as the Orbitrap Tribrid Fusion instrument, also used for this thesis work, routinely identify 6000 to 7000 phosphosites per sample from non-stimulated samples (Ferries et al., 2017). Advancement in phosphoproteomic technology and increased interest in understanding phosphorylation, especially in cancer, has led to the identification of 200,000 human phosphorylation sites (Hornbeck et al., 2015). The problem, however, lies in knowing the functional relevance of these phosphosites. To address this issue,

a recent study manually curated 112 publicly available datasets of phospho-enriched proteins creating a reference phosphoproteome containing 119,809 phosphorylation sites and through machine learning these phosphosites were prioritized according to a predicted functional score as an indication of functional relevance (Ochoa et al., 2019). In the context of CHO cells, our group published the first global comparative phosphoproteomic analysis of CHO-K1 cells expressing human secreted alkaline (SEAP) subjected to temperature shift and used the combination of both IMAC (Fe-NTA) and MOAC (TiO_2) phosphopeptide enrichment (Henry, Power, et al., 2017a). Using an LTQ Orbitrap XL mass spectrometer the total number of phosphopeptides identified was 2466 from the TiO_2 enriched samples and 3351 from the Fe-NTA enriched samples. In a preliminary study of CHO cell growth curve, I also used a combination of Fe-NTA and TiO_2 to enrich phosphopeptides from CHO cells collected every day from a four-day cell culture (Day1, Day2, Day3, and Day4). Figure 1.3 shows a bar graph of all phosphopeptides and phosphosite identified from this study using both enrichment techniques and the same mass spectrometer as used in Henry et al. (2017). Due to the high number of phosphosites identified by $TiO_{2,}$ a decision was made to use the TiO_2 enrichment technique for phosphopeptide enrichment presented in Chapter 2, Chapter 3 and Chapter 4 of this thesis.
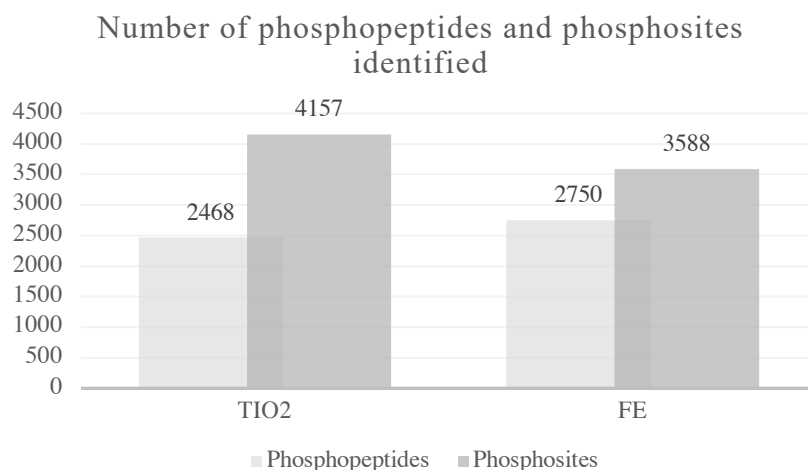
Figure 1.3 Qualitative analysis of phosphopeptides and phosphosites identification. Graph showing total phosphopeptide and phosphosite identifications by TiO₂ and Fe_NTA (FE) chromatography of CHO DP12 whole cell lysate by LC-MS analysis

1.4.6 Phospho-tyrosine enrichment.

A limitation of chromatography-based phosphopeptide enrichment techniques that are also used in this thesis work is the under-representation of phospho-tyrosine residues. This under-representation is partly due to the sub stochiometric of phosphorylated tyrosine (1%) compared to phosphorylation on Threonine (9%) or Serine (90%) residues, as shown in Figure 4.2. To increase the number of tyrosine residues, phosphopeptide enrichment is often performed by the use of immune affinity enrichment techniques using anti-phospho-tyrosine antibodies (Lombardi et al., 2015).

1.4.7 Identification of peptides by mass spectrometry

A mass spectrometer is designed to separate gas-phase ions (generated by an ion source) according to their m/z (mass to charge ratio) under the influence of a magnetic or an electric field (in a mass analyzer), and then record the abundance of each ion type in a detector. Peptides/protein samples are converted into gaseous phase ions with the help of an ion source (see Figure 1.4). Depending on the nature of the analysis and sample properties, an ion source is carefully selected to couple with a compatible mass analyzer. Two dominant classes of ion source used in proteomic sample analysis are

Matrix-Assisted Laser Desorption Ionization (MALDI) (Karas & Hillenkamp, 1988) and Electrospray Ionization (ESI) (Fenn at al., 1989).
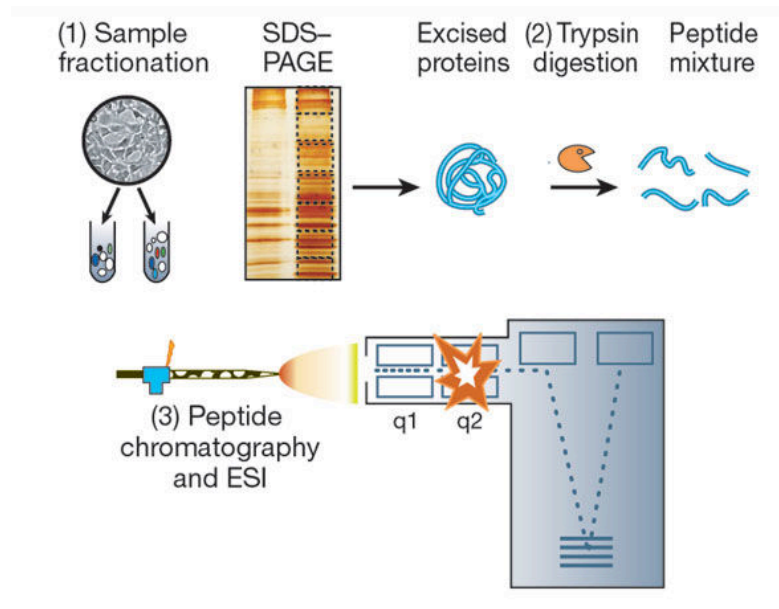
Figure 1.4 A schematic illustration of mass spectrometry-based proteomics workflow. (1) Samples are often fractionated before digestion to decrease the complexity of the proteomic sample. (2) In the bottom-up approach, proteins are digested into smaller peptides fragments before (3) injecting into a mass analyser using an ion source.

In the case of MALDI, samples are spotted onto a metallic plate by co-crystallization with a solid matrix such as CHCA or Sinapinic Acid. Samples are then ionised by pulsed laser irradiation in a vacuum, and the matrix ionises the analyte "A" by acting either as a proton donor $[A+H]^+$ or a proton acceptor $[A-H]^-$ (J. K. Lewis, Wei, & Siuzdak, 2000). Due to the high amount of chemical noise created by the ionizing matrix, MALDI is not suitable for analysis of low molecular weight compounds or peptides. However, the technique is widely used for structural characterisation of large protein molecules in top-down proteomics (Hrabák, Chudácková, & Walková, 2013). In ESI, ions are generated by passing the liquid phase peptide sample through a narrow capillary (internal diameter < 250 μm) at atmospheric pressure. An electric potential difference of +500 to +4,500V is applied relative to a counter electrode. The passage

of the analyte solution through the narrow orifice of the capillary creates a fine mist of solvent with peptide mixture in a "Taylor cone" architecture. Often the spray is assisted by a coaxial gas flow such as nitrogen. The droplets emitted from the Taylor cone are positively charged (positive ion mode) and undergo rapid evaporation. A shrinking droplet finally transfers charge from the solvent outer surface to the analyte in a process termed as "Coloumbic repulsion" (Marginean, Nemes, Parvin, & Vertes, 2006). These droplets are pulled inward into the mass analyser by the instroment's internal vacuum as illustrated in Figure 1.5.
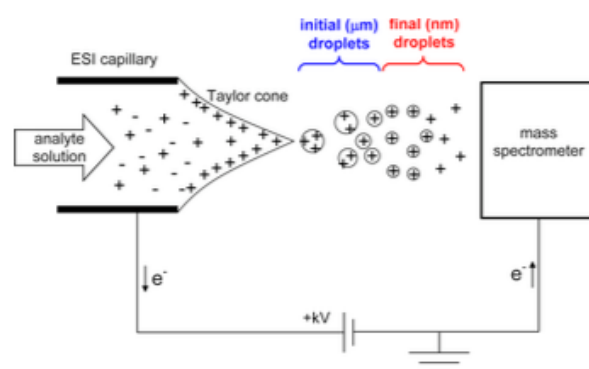
Figure 1.5 Schematic depiction of an ESI source operated in positive ion mode (Konermann, Ahadi, Rodriguez, & Vahidi, 2013). From left to right, the volatile solution containing analyte molecules leave the electrically potentiated ESI capillary. Solvent evaporation reduces the droplet size, ultimately leaving the positively charged analyte to enter the mass spectrometer for detection.

A significant advantage of ESI over MALDI it that it is a soft ionisation technique that retains labile PTMs such as phosphorylation for analysis by the mass analyser. ESI also provides the ability to couple MS with liquid separation technique such as nanoflow UHPLC to allow a rapid and thorough investigation of complex biological samples and is known as a nano-electrospray ionization (nano-ESI) (C. S. Ho et al., 2003). Typically used for peptide and protein analysis, nano-ESI allow proteomic investigations from microlitres of the sample by maintaining a flow rate of nL/min at which samples are introduced into the mass spectrometer. Smaller droplets created by the nano-ESI allow higher ionization of analyte molecules emitting from the needle

tip and has been shown to reduce interference from salt and other contaminating species thus providing better sensitivity (Faramawy et al., 2005).

Proteomic samples discussed in this thesis were resolved in a C-18 column under organic solvent gradient prior to injection into the mass spectrometer by nano-ESI. Typically, 1 µg (1 µL-2µL) equivalent of peptide mixture was picked up by the robotic auto-sampler for each LC-MS run. The separation of a peptide mixture under acetonitrile (ACN) gradient allows sequential elution of peptides according to their hydrophobicity thus preventing sample overload in the mass spectrometer and increases the sample acquisition time allowing for more scans per sample.

1.4.8   Mass analysers

Different types of mass analysers measure the mass of ions in different ways, and each has its own list of pros and cons. The fundamental difference between the mass analysers lies in precision and accuracy. The precision of an instrument is the resolving limit of two adjacent peaks, whereas the accuracy of a mass analyser is the measure of correct determination of the ion mass (Mann & Kelleher, 2008).

Mass analysers can be divided into two broad categories: a beam type mass analyser and a trapping mass analyser. In a beam type mass analyser, ions leave the ion source as a beam and travel through the analysis field towards the detector, for example, a time of flight (TOF) mass analyser, which as the name suggests measures the time taken by the ions to reach the detector. TOF instruments offer the unlimited mass range at a very reasonable cost but are limited in their ion filtering capabilities, therefore, are suitable for less complicated samples to avoid multiple ions reaching the detector at the same time (Mamyrin, 2001). In a trapping type, mass analyser ions are trapped inside the analysis field, and the frequency of analysis field is used to manipulate the path of oscillating ions, thus stabilizing selected ions only. An example of this system

is the Orbitrap-type mass analyser. A spindle-like central electrode is enclosed within two electrically independent concave-shaped outer electrodes; the thin gap between external electrodes and central electrodes acts as an analysing field (Makarov et al., 2006). The ions are kept in a complex spiral motion by the correct application of a voltage between the outer and the inner electrodes. Trapped ions generate an image current specific to their mass-to-charge ratio, which can be detected and converted into the mass spectrum by Fourier transformation of the frequency signal. Other types of trapping type mass analysers include a quadrupole, ion trap and FTMS type mass analyser. A quadrupole contains 4 parallel conducting rods paired at RF frequency, 180 out of phase. Alternating the RF frequency causes ions to take spiral trajectory, the stability of which depends on ion m/z ratio; therefore by manipulating the RF frequency of a quadrupole selected ions are stabilized and others are thrown out of the analyser. Quadrupole type mass analysers are therefore ideal for ion filtering and offer very high scan speeds but offer a limited mass resolution. The ion trap mass analyser operates like a quadrupole but instead of trapping ions between linear rods, ion trapping accrues in a central cavity inside a ring electrode with two hyperbolic, electrically common, end cap electrodes. The ring electrode is operated by fixed RF and the end cap electrodes are maintained at oscillating AC potential. A combination of RF and AC potential enables stabilization of all ions within a selected m/z range within the ion trap. By manipulating the AC potential of the end cap electrodes, unwanted ions can be ejected from the trap. The ion trap allows trapping of single-parent ions that can then be fragmented into daughter ions, which are also trapped and measured by the ion trap (MS/MS), an instrumental technique for sequence identification of peptides. Many mass spectrometers in the market provide a combination of mass analyzers to

complement the cones of other mass analyzers. The Thermo Scientific LTQ XL mass spectrometer, previously mentioned, offers a combination of a linear ion trap and Orbitrap mass analyser (Andrews, Dean, Hawkridge, & Muddiman, 2011). Data presented in this thesis from chapter 2 to chapter 4 was generated using the newer state-of-the-art Orbitrap Fusion™ Tribrid™ Mass Spectrometer, that offers a combination of Quadrupole mass filter, an Ultra-High-Field Orbitrap mass analyser and Duel-pressure Linear ion trap (Ferries et al., 2017). This instrument can routinely identify 3-4 fold more confident phosphopeptide identifications from the same about amount of sample than was used for analysis using the older Orbitrap XL instrument.

1.4.9 Tandem mass spectrometry

A single scan of peptides in the mass analyser provides the mass of intact peptides. However, no information about the sequence of the peptide is obtained. Tandem MS/MS strategies select precursor ions based on the first MS scan, and these chosen ions are isolated (parent-ions) in an approach called data-dependent acquisition (DDA). Parent-ions are then subjected to disintegration by collision with a chemically inert gas such as Helium or Argon to release product-ions. A round of second scan (MS/MS) is then used to measure the mass of fragmented product-ions. Some instruments are capable of repeating this process, leading to the fragmentation of already fragmented ions in what is called an MS^n experiment. The MS^n technique is particularly useful to obtain sequence information of a phosphopeptide. The presence of a phosphate residue on a peptide shields it from complete fragmentation, and also the first collision often leads to the neutral loss of phosphoric acid which dominates the scan survey peak obscuring the partially fragmented product-ions. A second fragmentation event is used for complete fragmentation of the peptide. The mass of product-ions obtained from the second fragmentation event is measured by a third

scan event (MS^3) (Potel, Lemeer, & Heck, 2019). Selected parent ions can be fragmented in various ways such as by Collison Induced Dissociation (CID), Electron Transfer Dissociation (ETD), and Electron Capture Dissociation (ECD).

1.1.1.1 Collision induced dissociation

Fragmentation by CID is achieved by the transfer of kinetic energy from accelerated inter gas (He, N_2, or Ar) molecules to the polypeptide. CID cleaves the amide bond among the peptide backbone resulting in the formation of two types of product-ions; ions containing the N-terminal (called b ions) and the other type containing the C terminal (y ions). A peptide sequence is reconstituted by creating a ladder made up of b ions and y ions. A ladder sequence is deduced by calculating the difference in the mass of the peptide fragments as they increase in size from N-terminus or C-terminus, as shown in Figure 1.6. A manual interpretation of the fragmentation spectrum can reveal sequence information from a peptide spectrum; however, manually obtaining sequence information of all the peptides present in a whole-cell lysate would be impractical. Computational search algorithms such as SEQUEST or MASCOT are used to search peptide ions against publicly available databases to identify the best sequence match between data obtained from the mass spectrometer and the database (Wright et al., 2012). Figure 1.6 shows an example of sequence fragmentation by CID and the resultant product-ions.
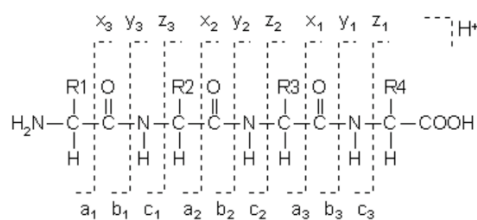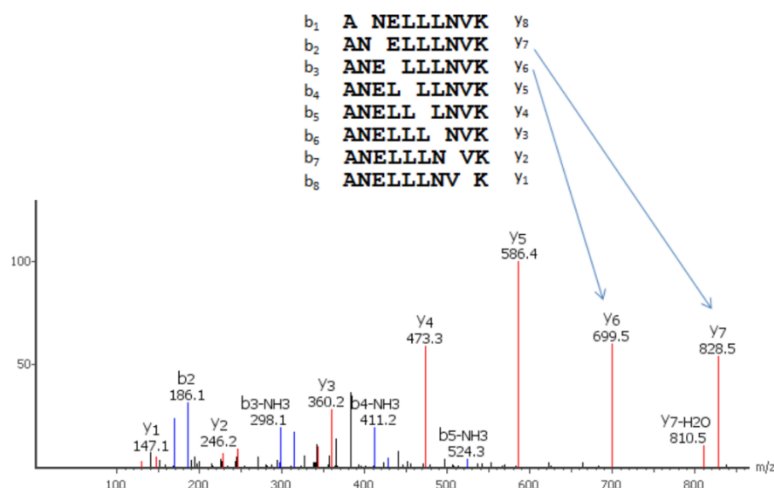
Figure 1.6 Top) A peptide backbone showing theoretical fragmentation sites and resultant ion type formation. Bottom) Fragmentation pattern of example peptide ANELLLNVK and the ion ladder made up of b-ions and y-ions resulting from its fragmentation.

1.1.1.2 Higher Energy Collisional Dissociation

Higher Energy Collisional Dissociation (HCD) is essentially CID when fragmentation is performed in a dedicated collision cell. The product-ions are then transferred to the Orbitrap for high-resolution detection via the C-trap. The term "higher energy" in HCD refers to the higher radiofrequency applied to the C-trap to retain fragment ions (Olsen et al., 2007). Fragmentation detection by the Orbitrap offers higher confidence in structural identification reflected by a higher search score as compared to CID; however, at a reduced scan speed due to increased cycle time (Jedrychowski et al., 2011).

1.1.1.3 Electron transfer dissociation

As the name of the technique suggests, fragmentation by ETD is achieved by the transfer of an electron from a pre ionized ion to the peptide carrying multiple positive charges, thus creating an unstable peptide cation radical. The unstable peptide cation ultimately breaks down between the alpha-carbon atom and nitrogen atom resulting in the formation of c and z ions. Fragmentation by ETD is particularly useful where preservation of a side chain or a chemical modification on a peptide needs to be

preserved, for example, phosphorylation or glycosylation (M. S. Kim & Pandey, 2012) (Riley & Coon, 2018).

1.1.1.4   Hybrid fragmentation strategies

Newer state-of-the-art instruments offer multiple fragmentation strategies such as EThcD offered on the Orbitrap Fusion Tribrid instrument (used in this thesis). In EThcD, initial fragmentation is performed by ETD, which preserves the labile modifications such as phosphorylation and generating c/z ions followed by HCD fragmentation creating b/y ions resulting in more abundant spectra compared to ETD or HCD alone (Frese et al., 2013).

1.4.10 Identification of phosphopeptides by mass spectrometry.

The phosphate group attached to a peptide is relatively labile and competes with the peptide backbone during CID fragmentation if the peptide is selected by first scan (MS^1). Energy transfer by CID results in neutral loss or H_3PO_4 (98 Da) or HPO_3 (80 Da), and the resultant spectrum is dominated by a peak at 98 Da or 80 Da less than the precursor ions mass (DeGnore & Qin, 1998) from the second scan (MS^2). The detection of a neutral loss triggers a second fragmentation event to obtain structural information from the phosphosite bearing peptide inform a third scan event (MS^3). The additional scan cycles required for MS^3 spectra reduces the number of scans that can be performed in a given instrument time. A new method used in this thesis called Multistage Activation (MSA) or "pseudo-MS^3" circumvents the MS^3 problem by assuming that all MS^1 spectrums contain a neutral loss, and the neutral loss triggers MS^2 while the precursor ion is still present in the chamber thus avoiding the ion isolation step (Erickson et al., 2015). Figure 1.7 shows the schematics of MSA spectra acquisition compared to Neutral loss MS^3 Spectra. An example of MSA fragmentation spectra obtained from the sample used in the thesis is shown in Figure 1.8.
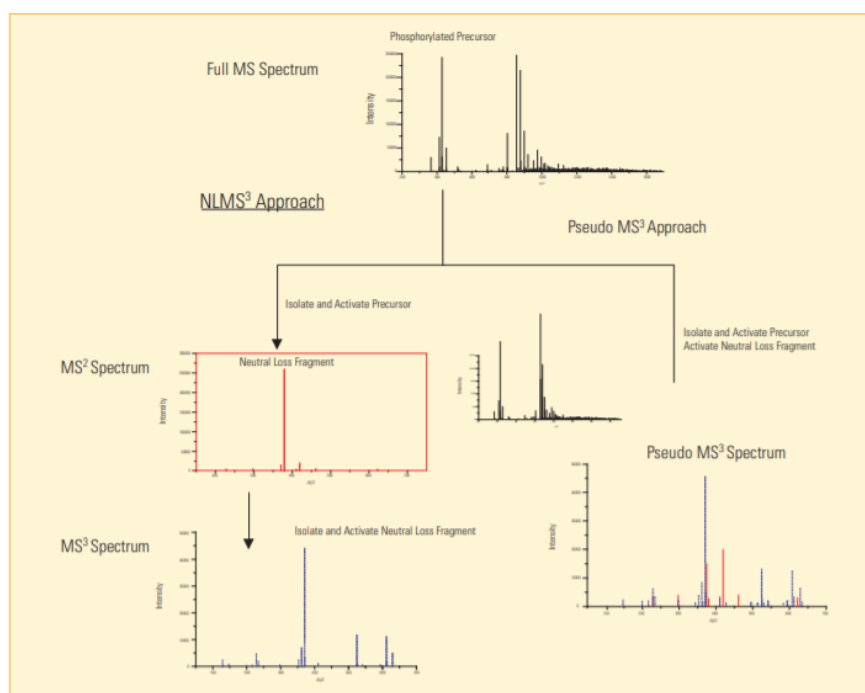
Figure 1.7. Schematics of Multistage activation. For Neutral loss MS^3 (Left), when the MS^2 spectrum is dominated by the neutral loss fragment, MS^3 is triggered on isolated neutral loss fragment. Multistage activation (right), neutral loss ion is fragmented in the presence of precursor ion resulting in more abundant spectra.
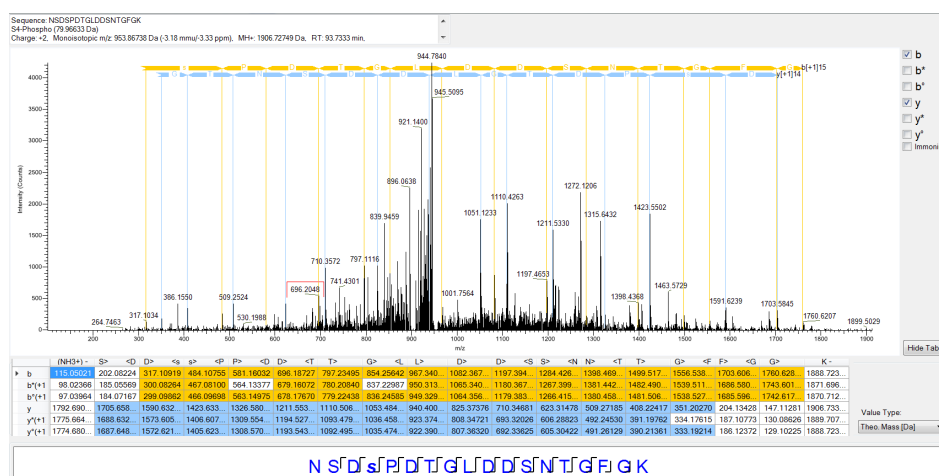
Figure 1.8 Multi-Stage Activation Spectra from phosphopeptide NSDS*PDTGLDDSNTGFGK from a sample used in chapter 4. b-ions and y-ions generated from the CID fragmentation are shown in yellow and blue, respectively.

Several software tools are designed to exploit this fragmentation information to localise the site of phosphorylation. Search engines such as MASCOT (D N Perkins, Pappin, Creasy, & Cottrell, 1999) and SEQUEST (Eng, Mccormack, & Yates, 1994) provide a "delta score" for localisation and identification of all possible phosphosites. While MASCOT and SEQUEST are excellent tools for protein identification, they are

not optimised for PTM analysis. A number of tools dedicated exclusively to PTM analysis for MS/MS spectra have been developed. Tools like Ascore (Beausoleil et al., 2006) and PTM score (Olsen et al., 2006) use probabilistic-based approaches by generating a theoretical list all possible phosphorylation sites for the peptides identified from the database and then by matching this information with experimental MS/MS spectra for the presence or the absence of site-determining fragment ions. A widely used tool PhosphoRS also utilises this probabilistic-based approach to assign site localization. The unique feature of PhosphoRS is its ability to provide information based on multiple fragmentation methods, including CID and ETD. For the work presented in this thesis, only the phosphopeptides with PhosphoRS phosphosite probability higher than 75% are considered as confident phosphosite identifications as suggested and used by the PhosphoRS developers (Potel et al., 2019).

1.5 Quantitative proteomics

With advances in mass spectrometric-based proteomics, the number of confident identifications has increased, and to analyse the complex biological data came the necessity to develop computational and practical tools to analyse and statistically evaluate data from LC-MS experiments. Quantitative techniques provide an opportunity to compare the signal from the same peptide or protein but from different experimental conditions. Various methods have been developed to obtain statistically significant quantitative data from LC-MS proteomics experiments. Traditionally, quantitative strategies for LC-MS analysis have relied on sample labelling to quantify quantitative data accurately. Labelling techniques such as iTRAQ (Isobaric labelling) (Ross et al., 2004), ICAT (chemical labelling) (Gygi et al., 1999), and SILAC (metabolic labelling) (Roux, Lison, Junot, & Heilier, 2011) have been used with great success in large-scale proteomic experiments. Different labelling techniques have

their advantages and disadvantages. Overall, labelling techniques provide a substantial reduction in quantitation errors and analysis time by minimizing errors introduced by sample preparation and by allowing sample pooling and multiplexing. These methods, however, require meticulous sample preparation and expensive labels. Moreover, only as many samples can be analysed as there are different labels available which are probably the most significant limitation of quantification by sample labelling. The label-free approach overcomes these barriers. This is the method of choice used in this thesis, in chapters 2-4.

1.5.1 Label-free quantitation

As suggested by the name, label-free quantification techniques do not require additional sample manipulation steps. The method exploits the peak intensity of peptide ions or identification frequency for a particular peptide to obtain quantitative data. Unlike labelled approaches, where only modified isotope identification leads to quantification, in a label-free approach any peptide identified could contribute to quantitative data, which means that theoretically, there is no limit to the number of different samples that can be analyzed by a label-free approach (Nikolov, Schmidt, & Urlaub, 2012). Label-free quantification techniques can be distinguished into two main groups: (A) measuring and comparing the signal intensity of peptide precursor ion, called area under the curve approach (AUC) or by (B) counting the fragment spectra of identifying peptide of a given protein, referred to as spectral counting (G. Zhang et al., 2010).

In the spectral counting approach relative quantification is achieved by counting the number of tandem MS spectra of identified peptides for a particular protein and then comparing the number of such spectra among different experiments (Lundgren, Hwang, Wu, & Han, 2010).

For the ion intensity technique, the ion chromatograms for each peptide are extracted from an LC-MS run, and their mass spectrometric peak areas are integrated over the chromatographic time scale (Bondarenko, Chelius, & Shaler, 2002). Quantitative information is then obtained by comparing intensity values for each peptide compared to their respective signals in one or more other different LC-MS runs (Higgs et al., 2005). The analytical accuracy of quantification by ion intensity depends mainly on chromatographic reproducibility among different samples because mismatched chromatographic profiles could lead to a comparison between different ions eluting at the same time (Ong & Mann, 2005). Software dedicated to computational analysis of ion intensity quantitative experiments such as the Progenesis QI for Proteomics from Nonlinear Dynamics (http://www.nonlinear.com/progenesis/qi/) are designed to adjust mismatch between runs by automatic adjusting or manually dragging the chromatogram along the time axis to find the best fit (Bylund et al., 2002). This software is used for the differential proteomic and phosphoproteomic analysis as outlined in chapters 2-4. Progenesis QI for Proteomics software is essentially an image analysis software that performs the LC-MS run alignment by selecting a reference set to which all other runs are compared. An aggregate data set from all aligned runs containing all peak information is created and matched with each sample to provide differential information (http://www.nonlinear.com). A schematics of Progenesis QI for Proteomics analysis for differential peptide expression identification is shown in Figure 1.9.
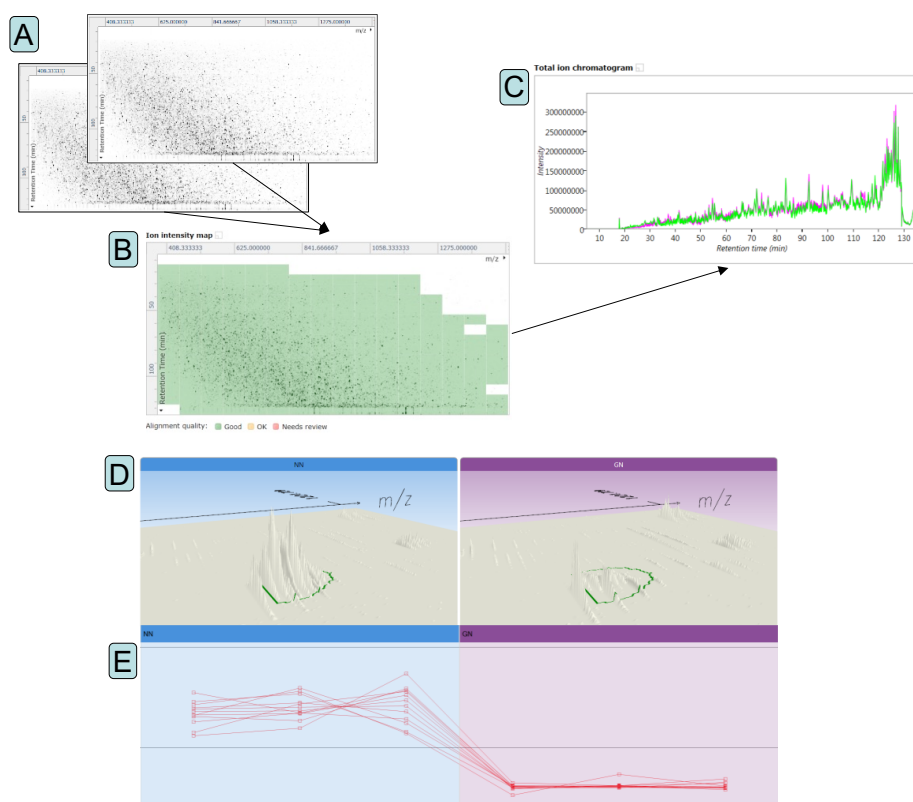
Figure 1.9 Workflow of differential expression analysis by Progenesis QI for Proteomics. A) Each LC-MS run is visualised as a 2D image by spreading detected m/z values of eluting peptides on the x-axis over sampling time on the y-axis. B) Images are overlaid, and areas with poor overlap are identified. C) automated sample alignment feature adjusts the chromatogram by sliding the pictures along the time axis. D) A 3D representation of a selected peptide peak between two experimental conditions. E) Expression profile of all contributing peptides from protein GLUL between GN and NN cell line used in chapter 4.

To demonstrate the effect of sample preparation on the quality of label-free quantification, Table 1.1 compares the Progenesis QI software alignment score percentage of samples processed using single trypsin digestion and a trypsin-trypsin double digestion methodology. As can be seen in Table 1.1 the average alignment score percentage for the single digested samples is 64 % whereas the same samples, when double digested by trypsin-trypsin, provides an average alignment score percentage of 96% indicating improved run to run reproducibility.

Progenesis QI Alignment Score		
Samples	Single digestion	Double digestion
Sample 1_Rep1	50	95
sample 1_Rep2	80	98
sample 1_Rep3	56	Ref
sample 2_Rep1	66	96
sample 2_Rep2	Ref	98
sample 2_Rep3	67	98
sample 3_Rep1	85	97
sample 3_Rep2	65	97
sample 3_Rep3	44	94

Table 1.1 Alignment score percentage from Progenesis QI for Proteomics software from three biological replicates of CHO DP12 cells. Each biological replicate sample was analysed three times as technical replicates. The alignment score was significantly improved after double digestion of protein samples. Ref: - reference run to which all other runs are aligned.

1.5.2 Protein identification

The mass spectrometer output provides spectral information about precursor ions and mass of amino acids sequence in a raw file format. This spectral information must be searched against a theoretical protein sequence database often derived from genome sequencing. The spectral data from the raw file is interpreted by software tools and converted into peptide sequences under a set of user-defined conditions such as mass tolerance for precursor and fragment ion, type of protease used, number of tolerated missed cleavages, and the presence of modifications such as phosphorylation. The confidence in a peptide sequence is reflected by its peptide score. These peptide sequences that pass the confidence threshold are then matched against protein databases to identify parent proteins from which the sequence is likely to have come from thus providing a score on the likelihood of a particular protein is present in the original sample. Two popular search engines that match mass spectrometry data to

protein databases are MASCOT (Perkins et al., 1999) and SEQUEST (Eng et al., 1994). The quality of mass spectrometric based proteomic investigations is inherently affected by the quality of the sequence database against which the peptides sequences are searched. This is of particular significance to CHO cell proteomics because in the past large-scale CHO cells proteomics was hindered by the non-availability of a reference CHO genome from which protein sequences could be derived. Therefore, spectral files obtained from MS analysis of CHO samples were searched against human/rat or mouse databases to search for homologues peptide sequences to infer the presence of conserved CHO proteins. The first breakthrough in CHO cell sequencing happened with the publication of the CHO-K1 genomic sequence in 2011 when 24,883 predicted genes were made publicly available through CHOgenome.org (Xu et al. 2011). This is outlined in more detail in section 1.8.

1.6   Functional annotation

Advancements in mass spectrometric based qualitative and quantitative techniques have allowed investigation of large-scale data, the results for which are mostly generated as a long list of protein identifications, usually containing a protein name and accession number. However, little or no biological information is provided. Depending on the context in which the experiments are performed, higher-level biological details are usually obtained by searching a protein list against external databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) which contain functional information about genes or protein collected over time from experimental validation or predictions. Gene Ontology Consortium (GO; http://www.geneontology.org) has provided a comprehensive resource of evidence-supported annotations to describe the biological role of proteins by classifying them into ontological categories: biological process, molecular function and cellular

component (Gene Ontology Consortium, 2004). Databases such as KEGG and GO resources are often integrated by web tools such as Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/) to provide comprehensive functional analysis of proteins in one search effort (D. W. Huang et al., 2007). Databases such as STRING (string-db.org) contain publicly available information on protein-protein interactions and enrich these with computational prediction to generate protein interaction maps from the list of proteins searched (Szklarczyk et al., 2019). Most of these resources arise from publicly funded research; therefore, they are free to use. Commercial software such as Ingenuity Pathway Analysis (IPA) has dedicated teams of bioinformatics scientists which routinely curate the database information and are therefore more reliable and can even predict the activation of downstream molecules from the differential genomic data (Krämer, Green, Pollard, & Tugendreich, 2014). An example of this approach can be seen in Chapter 3 of this thesis where the activation status of transcription factors was predicted from differential transcriptomic data between a low productivity CHO cell line and a high productivity CHO cell line.

1.7   Chinese hamster ovary cells

Unlike small pharmaceutical molecules which can be chemically synthesised biopharmaceuticals are therapeutic protein molecules of biological origin. These are large and heterogeneous proteins that can only be manufactured in living host systems by incorporating a genetic sequence that codes for the protein of interest into the genome of a host cell by recombinant DNA technology. Figure 1.10 shows the relative scale of complexity and size between a well know chemical-pharmaceutical Aspirin, Somatotropin hormone produced in recombinant yeast, and an iconic monoclonal antibody, Herceptin, which is produced in recombinant Chinese hamster ovary cells. Parallels are drawn for similar size and scale from the automotive industry by McKinsey & Company.

Size of 3 well-known pharmaceuticals

Aspirin, 21 atoms

| Somatropine, ~3,000 atoms | Herceptin, ~25,000 atoms |

Comparable objects[1]

Bike, ~20 lb

| Car, ~3,000 lb | Business jet, ~30,000 lb (without fuel) |

Relative cost of development and complexity

Figure 1.10 A visual representation of the relative complexity and size of the different class of pharmaceutical products (blue) compared to the difference in size and complexity in the automotive industry. Source: https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/rapid-growth-in-biopharma#

The choice of a host system primarily depends on its ability to produce a biologically active product, which involves correct folding of the synthesised protein and assigning of suitable posttranslational modifications. Logically, for producing therapeutic

35

proteins for human use, the obvious choice of host system appears to be a human-derived cell line. However, to provide therapeutic protein products at an industrial scale, an ideal host system should also offer sustainable industrial characteristics such as robust growth, adaptability, ability to grow in chemically defined media in suspension culture, and high specific productivity of the protein product. Chinese hamster ovary (CHO) cells offer these technical characteristics which most human cell lines lack (Dumont, Euwart, Mei, Estes, & Kshirsagar, 2016). CHO cells are also able to correctly fold a recombinant protein and assign tolerable "human-like" post-translation modifications (Hansen, Pristovšek, Kildegaard, & Lee, 2017). Therefore, the biopharmaceutical industry has used CHO cells as a preferred host system for industrial production of recombinant protein products since 1987 when Genentech got FDA approval for Activase[TM], a tissue plasminogen activator, as the first therapeutic molecule produced in CHO cells (Jayapal et al. 2007). Currently, monoclonal antibodies (mAb) dominate the biopharmaceutical landscape and occupy 65% of the biopharmaceutical market by sales. Between the year 2015-2018, 68 new therapeutic mAb products were approved, and 57 of them (85%) are produced in CHO cells (Walsh, 2018). Figure 1.11 shows the top 10 biologics sold worldwide in 2017, six of which are mAbs and five out of these six antibodies are manufactured in CHO cells (Urquhart, 2018).
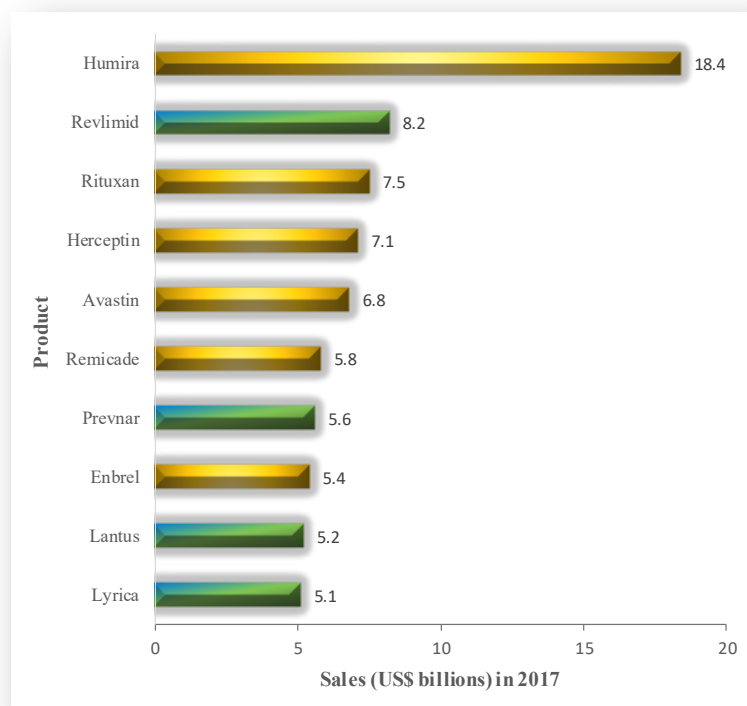
Figure 1.11 – The 10 top-selling biologics worldwide in 2017. In the bar graph adapted from Urquhart., 2018, the golden bars show therapeutic monoclonal antibodies. Humira, Rituxan, Herceptin, Avastin and Enbrel are produced in CHO cell lines.

As the number of new mAb product is likely to increase in the near future, increased demand for CHO cell lines that can sustain increased market needs will be required. Also, due to the increasing complexity of new innovative products (Tejwani et al., 2018) and with many top biopharmaceuticals that are currently being produced in CHO cells coming off patent in the next few years (Dahodwala & Sharfstein, 2017a), the need for high productivity CHO cell lines is bound to increase. The CHO cell research community has realised this trend, and the sustained efforts towards improving capabilities of these cells have led to remarkable improvements in their recombinant protein production capabilities. The first wave of development was made in bioprocess engineering optimisation and media design, but these approaches are reaching saturation (Hacker, De Jesus, & Wurm, 2009). The second wave included targeted engineering of selected factors to improve productivity (Laux et al., 2013).

The third wave today is utilising systems-level information to engineer cellular pathways to improve bioprocess outcomes (Kuo et al., 2018). Fundamental to this approach is a detailed understanding of CHO cell biology, especially from large scale omics studies in the bioprocess context to enable CHO researchers to identify and characterise cellular mechanisms at a systems level for manipulation via combinatorial genome editing techniques (Lee, Grav, Lewis, & Kildegaard, 2015) and computation modelling approaches (Sha, Huang, Wang, & Yoon, 2018) to achieve desirable bioprocess outcomes. To this end, CHO cell research has mainly focused on understanding the genomic, transcriptomic, and to some extent, proteomic level information in CHO cells (Yusufi et al., 2017). Other regulatory mechanisms such as cellular regulation by reversible post-translational modification of cellular proteins such as phosphorylation have lagged behind. Through the discussion in the following sections, this thesis identified this knowledge gap and provided a rationale for the application of phosphoproteomics in CHO cell research. Subsequent chapters address this gap by studying three fundamentally important bioprocesses attributes of CHO cell using large-scale, quantitative phosphoproteomics.

1.8 Omics approach towards CHO cell engineering

The current status of CHO cell proteomics and phosphoproteomics depends on developments in CHO genomics and transcriptomics. As discussed previously in section 1.5, mass spectrometry-based shotgun proteomics relies on the availability of a protein sequence database against which the spectral data is searched. These databases are often derived from genomic sequencing. The sequencing of the CHO-K1 genome achieved a significant milestone in CHO cell research in 2011, which predicted 24,383 CHO genes (Xu et al., 2011). Later in the same year, the result of two other CHO sequencing projects was published; 1) Sequencing of CHO-SEAP

cells by Illumina sequencing (Hammond et al., 2011), and 2) cDNA sequencing of CHO-K1 and CHO DUKXB11 cell by pyrosequencing technology (J. Becker et al., 2011). Using the sequencing information from Xu et al.(2011) and Becker et al. (2011), our group showed a 40 – 50% increase in protein identification along with an increase in protein confidence as compared to databases derived from *Mus musculus* and *Rattus norvegicus* (Meleady et al., 2012). Since the immortalisation of the CHO cell line in 1956 by Theodore Puck (Puck et al., 1958), CHO cells have diverged into multiple lineages as shown in Figure 1.12 (Xu et al., 2017). Each family has undergone extensive mutagenesis and clonal selection (Wurm & Hacker, 2011). Hence, to account for genetic variability among CHO cell lines and to create a representative genomic sequence of all CHO genes, Lewis et al. (2013) sequenced six CHO cell lines from the CHO-K1, DG44 and CHO-S lineages along with the sequencing of Chinese hamster tissue (N. E. Lewis et al., 2013b). Sequence information from the 2013 hamster genome and the original CHO-K1 genome by Xu et al. (2011) was annotated and released by NCBI RefSeq in 2012 and 2014 respectively.
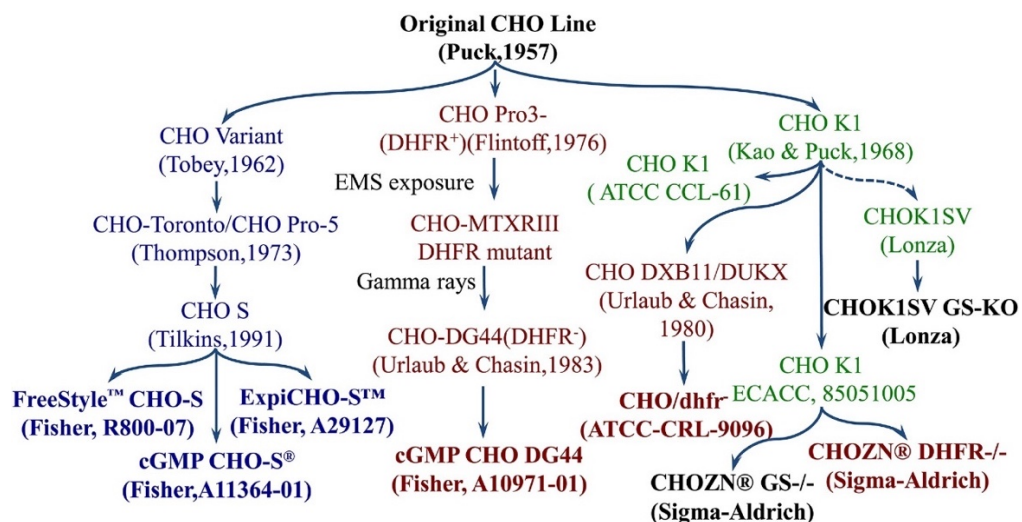
Figure 1.12 Cell lineage of CHO cells. The figure is taken from N. Xu et al. (2017)

More recently, the Chinese hamster genome was re-sequenced using single-molecule real-time sequencing and Illumina-based assemblies (Rupp et al. 2018). This project

significantly improved the coverage of noncoding regions by covering >95% gaps from the previous draft genome. This genome was further refined by the Lewis group in June 2019 by a using proteogenomic approach to integrate proteomic, genomic and Ribo-Seq data to verify protein-coding regions in CHO cell line variants and identified 3529 new proteins compared to Refseq (S. Li et al., 2019). I prepared the proteomic mass spectrometry data used in this project in NICB. Finally, it is worth mentioning that to unify all sequencing efforts and to make available the CHO genomic information to the international research community, an open-source platform, CHOgenome.org was launched in 2011 (Kremkow, Baik, MacDonald, & Lee, 2015). These large scale omics efforts have enabled systems biology level analysis of large scale genomic (Hefzi et al., 2016), transcriptomic, (Lund et al., 2017) and proteomics (Park et al., 2017) data.

Reversible phosphorylation of proteins regulates virtually all processes in a biological system. Hence, it is valid to assume that bioprocess characteristics of CHO cells such as growth, recombinant protein productivity, cell death and apoptosis will also be affected by changes in protein phosphorylation levels. Moreover in the present era of multi-omic studies, incorporating a phosphoproteomic level of regulation to genome-scale models along with genomic and metabolic networks have the potential to improve our understanding of critical cellular networks (Gutierrez & Lewis, 2015).

1.9 Phosphoproteomics in CHO cells – Past and Present

Despite the importance of phosphorylation in cell metabolism and regulation (Ardito et al., 2017), it is surprising to see that global CHO phosphoproteomic studies have only recently started to emerge. On the other hand, phosphoproteomics has revolutionized the way cancer is targeted, as about 150 kinase inhibitor drugs are currently in clinical phase trials (Bhullar et al., 2018) and phosphoproteomics is

expected to play a significant role also in the next wave of cancer therapeutics, i.e., precision medicine (Ramroop et al., 2018) The earliest account of bioprocess related phosphorylation regulation in CHO cells found an alteration in the level of two phosphotyrosine proteins corresponding to 80kDa and 180kDa following temperature shift; an increase in the phosphorylation level of these two phosphoproteins (which were not identified by the authors of the study) was then hypothesised to play a crucial role in suppressed growth of CHO cells at low-temperature conditions (Kaufmann et al., 1999). Temperature reduction during the exponential phase of cell culture is a technique widely used by the biopharma industry to improve recombinant protein product yield (Bollati-Fogolín et al., 2005). Other studies have focused on understanding the behaviour of CHO cells through differential phosphorylation and have provided crucial information about the essential regulatory proteins and their post translations modification as a potential explanation for the increased protein product yield. Phosphorylation of eukaryotic initiation factor 2 (eIF2) has been linked with decreased recombinant protein production by CHO cells under stress-related conditions, suggesting the phosphorylation of eIF2 alpha as a potential CHO stress biomarker due to a direct relationship between eIF2 phosphorylation and protein synthesis rates (Underhill, Birch, Smales, & Naylor, 2005). This relationship between eIF2 phosphorylation and protein synthesis rates was further investigated in response to mild hypothermia by monitoring the changes in eIF2 total protein expression and phosphorylation levels. In this study, while the overall levels of eIF2 protein did not show a significant change upon temperature reduction, eIF2 phosphorylation levels were reduced at 32°C relative to phosphorylation levels observed at 37°C (Masterton et al., 2010). The cold shock response of eIF2 phosphorylation was subsequently linked to phosphorylation of eIF2A by specific kinase, eEF2K (eukaryotic elongation

factor 2 kinase). A reduced metabolic rate at low temperature gave rise to the activation of AMPK (5′AMP-activated protein kinase), which is upstream of eEF2K (Yamada et al., 2019). This group has also demonstrated the structural dynamics of heat shock factor protein-1 (HSF1) in relation to temperature shift in CHO cells. Hyper phosphorylation of HSF1 at high temperature causes HSF1 trimerization in the nucleus, which prompts its binding to heat shock elements, thereby activating the heat shock response (Roobol et al., 2009). This is an example of how phosphorylation plays a crucial role in activation, deactivation, and localisation of regulatory proteins in CHO. Another study linked phenotypic effects of mild hypothermia on CHO K1 cells to the activation of p53 by p38MAPK (p38 mitogen-activated protein kinase) (Tanel & Averill-Bates, 2007). The activation of p38MAPK was suggested to be linked with the activation of ATR (ataxia telangiectasia mutated- and Rad3-related kinase)-p53-p21 signalling pathway. The exposure of CHO-K1 cells to reduced temperatures leads to the activation of p53 by phosphorylation at Ser-15 through ATR activation. Simultaneous activation of p38 MAPK by ATR leads to phosphorylation of p53 at Ser-33 and Ser-46, which enhances Ser-15 phosphorylation. P53 subsequently activates downstream target p21, which leads to cell cycle arrest (Roobol et al., 2011).

Like ATR, other central signalling pathways have been investigated for their potential role in CHO cellular processes. In particular, the mTOR pathway, known as a central pathway for protein synthesis and growth regulation, was studied by engineering of a CHO cell line for constitutive expression of mTOR; the ectopic expression of mTOR resulted in increased viability, robustness, cell size, proliferation and antibody production by CHO cells (Dreesen & Fussenegger, 2011) suggesting mTOR as a potential cell engineering target. The role of the mTOR pathway was further investigated by chemical manipulation using rapamycin, a chemical inhibitor of the

mTORC1 protein. This study linked the phosphorylation status of eIF4E binding protein (4E-BP1) and S6 Kinase, two well studied downstream substrates of mTORC1 (Shah at al., 2000), as a measure of mTOR activity. Rapamycin addition on the seeding day led to growth arrest and higher specific productivity, and these changes correlated with de-phosphorylation of S6K1 at Thr-389, which is phosphorylated by mTOR. Rapamycin was found to differentially target S6K1 and 4E-BP1, which was shown by the increased phosphorylation of 4E-BP1 on day four and day six of culture after treatment with Rapamycin (Dadehbeigi & Dickson, 2015). A similar strategy was used to investigate the link between cell phenotypic behaviour and mTOR signalling, where mTOR activity was monitored throughout the batch culture of a panel of sister cell lines expressing different amounts of IgG4. In this study, the phosphorylation state of 4E-BP1 was shown to fluctuate throughout the culture, and eIF4E measures appeared to be co-regulated with the expression of recombinant protein (Josse et al., 2016).

With an increasing knowledge base of bioprocess-related cell signalling events and critical signalling molecules, these pathways are targeted to modulate CHO phenotypic behaviour to achieve desirable bioprocess attributes. For example, it is observed that even though high viable cell density (VCD) is the desired bioprocess attribute, uncontrolled proliferation could negatively impact culture outcomes. Hence, in order to control VCD to within an optimal range, cell cycle cyclin-dependent kinases (CDK) 4/6 were targeted by a small molecule inhibitor (Du et al., 2015). This study demonstrated an overall increase in specific productivity and no negative impact on product quality by using a small molecular inhibitor compound to achieve sustained G0/G1 arrest without impacting G2/M phase (Du et al., 2015). The studies, as mentioned above, show that a phosphoproteomic understanding of CHO cell biology

can provide an opportunity to target critical proteins and phosphoproteins for cell line engineering. However, such efforts have primarily remained limited, mainly because of lack of large scale, global phosphoproteomic studies in CHO cells. Thus far, only four studies have been published (3 of these involving our group) that utilizes mass spectrometry-based global quantitative phosphoproteomics to understand cell signalling events in CHO cells in bioprocess relevant contexts. Table 1.2 provides a reference to these studies, along with a short description of critical findings.

CHO phosphoproteomic studies	Highlight
1. (Henry, Power, et al., 2017a)	This study utilized a combination of TiO$_2$ and Fe-NTA chromatographic techniques and identified 700 differentially expressed phosphoproteins in IgG producing CHO-SEAP cells following temperature shift from 37 to 31 °C.
2. (Kaushik et al., 2018)	This work identified 1,415 differentially expressed phosphoproteins in IgG producing CHO DP12 cells at various phases of growth.
3. (Schelletter et al., 2019)	Combination of label-free and SILAC based quantative analysis identified varied response to IGF stimulation in different CHO clones.
4. (Dahodwala et al., 2019a)	A multi-omics study linked hyperactivation of transcription factor CREB1 in CHO cells with high productivity through interpretation of CHO transcriptomic, nuclear proteomic and nuclear phosphoproteomic data.
5. (Kaushik et al., 2019) *In review*	This study elucidated phosphoproteomic changes associated with adaptation of CHO-K1 cell line to growth in glutamine free media by identifying differential expression of 434 phosphoproteins between.

Table 1.2 Five global CHO phosphoproteomic studies published so far summarized.

During my PhD, I contributed to four of the studies outlined in Table 2 above. Three of these make the core research body of the thesis, namely; Kaushik et al., 2018 as Chapter 2, Dahodwala et al., 2019 as Chapter 3, and Kaushik et al., 2019 as Chapter 4. The first study of global CHO phosphoproteomics through mass-spectrometry published by our group which I contributed to involve the use of quantitative label-free LC-MS/MS phosphoproteomic analysis to gain insights into the molecular events associated with temperature reduction of CHO cell culture. Temperature shift conditions were created by reducing the culture temperature of a set of biological

replicates of CHO-SEAP cells to 31°C, 48 hours post-seeding. With the use of two phosphopeptide enrichment techniques, i.e., TiO_2 and Fe-NTA (IMAC) in combination with LC-MS/MS analysis by an LTQ Orbitrap XL™ Mass Spectrometer, a total of 700 phosphoproteins were found to be differentially expressed due to the temperature shift. Gene ontology analysis of these differentially expressed phosphoproteins showed that phosphoproteins that were increased in expression at low temperature were mainly involved in RNA processing and transcription factor activity whereas phosphoproteins that showed decreased expression at 31°C were mostly involved in chromosome maintenance and cell-cycleregulation, providing a potential explanation for the observed increased productivity and cell cycle arrest (Bedoya-López et al., 2016) at low-temperature conditions through phosphoproteomics (Henry, Power, et al., 2017b). In this study, I contributed to this study by validating the differential expression of two phosphoproteins NDRG1 and ATF2 by western blot analysis.

1.9.1   Preface to Chapter 2

Chapter 2 is a journal article that elucidates changes in the phosphoproteome of CHO cells at various phases of growth. Growth control strategies are often used in the biopharmaceutical industry to prolong the culture duration and increase protein productivity. As shown by Henry et al. (2017) a decrease in expression of phosphoproteins associated with cell cycle regulation in cells subjected to low culture conditions were also reflected by a reduction in doubling time of cells growing at 31°C (Henry, Power, et al., 2017b). Therefore to further investigate the role of phosphoproteins in CHO cell growth the phosphoproteomic dynamics of a four-day culture of IgG producing CHO DP 12 was investigated by quantitative label-free phosphoproteomic analysis, under the pretext that these changes may indicate cellular

state in terms of viability, growth and productivity, similar to mRNA and miRNA profiling of CHO cells during batch culture (Bort et al., 2012).

From a four-day CHO cell suspension culture, samples were collected each day from Day 1 to Day 4 to capture the exponential phase, stationary phase, and decline/death phase of culture. For phosphoproteomic analysis, phosphopeptides were enriched using Fe-NTA chromatography and analysed on an Orbitrap Fusion™ Tribrid™ Mass Spectrometer coupled to an Ultimate 3000 RSLC nano-LC system. To understand the progressive changes in the phosphoproteome from one day of the culture to the next, LC-MS/MS data files from the three replicate samples from two consecutive day samples were compared with each other, i.e., Day 2 versus Day 3, Day 3 versus Day 4, and Day 4 versus Day 5. The results from this project demonstrate the dynamic nature of the CHO cell phosphoproteome. Gene ontology analysis of differentially expressed phosphoproteins revealed 18 proteins associated with the mTOR signalling pathway, a pathway of profound interest in CHO research.

1.9.2 Preface to Chapter 3

Chapter 3 is a journal article that explores the molecular differences between an IgG producing CHO cell line (A0) and its progeny cell line (A1) by studying the transcriptomic, proteomic, and phosphoproteomic differences between these cells. The cell line model was generated by amplification of Transgene in A0 cell line by the stepwise increase in MTX concentration, achieving a 3-4 fold increase in transgene copy number and a resultant seven-fold increase in IgG productivity (Jiang, Huang, & Sharfstein, 2006). This study utilises a multi-omics approach to specifically target differential activation of the transcription factors in these two cell lines. Transcription Element Search System (TESS) that contains experimentally recorded binding consensus sequence predicted increased association of CMV promotor region with

transcription factors. The transcriptomic differences between A0 and A1 cells predicted the increased activity of CREB family transcription factors through Ingenuity pathway analysis. Although nuclear proteomic data showed a 1.3-fold less expression of CREB1 in high producing A1 cell line, the site-specific phosphorylation information obtained from the differential phosphoproteomic data suggested hyperactivation of CREB1 in the higher producer A1 cells. This observation was then validated by chromatin immunoprecipitation (ChIP). CREB1 transcription factor showed an ~6-fold increased association with the cytomegalovirus (CMV) promoter in the higher-producing cell lines.

1.9.3   Preface to Chapter 4

Chapter 4 is a journal article that explores the dynamics of protein and phosphoprotein expression in CHO-K1 cells in response to adaptation to growth in glutamine free media. Glutamine is supplemented to the CHO cell culture as a carbon source; however, the metabolic by-product of glutamine, ammonia, is toxic to the cells and it can also undesirably alter the recombinant protein product (Yang & Butler, 2000). Although various avenues for glutamine replacement (D. Y. Kim et al., 2013) and even removal (Wahrheit, Nicolae, & Heinzle, 2014) in CHO cells have been explored with reported increases in desirable quality traits, the responsible molecular mechanisms involved with these phenotypic changes are not fully understood. This study provides a comprehensive proteomic and phosphoproteomic landscape of CHO-K1 host cells adapted to growth in glutamine free media. Through the functional enrichment of 194 differential protein and 434 differentially expressed phosphoproteins, proteins associated with pre-mRNA signalling, Cytoskeleton rearrangement and Mapk1 pathway were identified.

2. Chapter 2 The Expression Pattern of the Phosphoproteome Is Significantly Changed During the Growth Phases of Recombinant CHO Cell Culture.

Dr Paula Meleady and I conceptualized the study. I performed the cell culture, collected and prepared sample for proteomic analysis. I carried out the formal data analysis, differential protein identification, data visualization, and discussion of the observed results. The resultant manuscript was drafted by me, which was reviewed and edited by Dr Paula Meleady.

Abstract

Post-translational modification of proteins by reversible phosphorylation plays a pivotal role in regulating key cellular processes, including transcription, translation, cell proliferation, differentiation, apoptosis, and signal transduction. Despite the importance of the phosphorylation level of regulation, little work has been carried out on the phosphoproteomic characterisation of Chinese hamster ovary (CHO) cells in bioprocess-relevant conditions. Growth control strategies are often used to prolong culture duration and increase specific productivity; however, the cellular mechanisms and regulatory pathways underlying growth strategies are poorly understood in CHO cells. Phosphorylation changes are dynamic and will respond to changes in culture conditions; this may reflect the status of the cells with respect to growth and viability of the culture. In this study, we have used a phosphopeptide enrichment strategy in conjunction with LC-MS/MS to carry out a large-scale differential phosphoproteomic analysis of IgG producing CHO DP12 cells at various phases of growth in serum-free suspension batch culture to characterise dynamic changes to the phosphoproteome with changing culture conditions. In total, over the various growth phases, we have identified 3,777 differentially expressed unique phosphopeptides from 1,415 differentially expressed unique phosphoproteins. Analysis of the whole cell lysate without phosphopeptide enrichment over the various growth phases revealed the differential expression of 834 unique proteins, with an overlap of 188 proteins between the proteomic and phosphoproteomic analyses. The inclusion of phosphoproteomic data significantly improves proteome coverage but also gives insights into the post-translational level of regulation during cellular growth of recombinant CHO cells.

2.1 Introduction

Chinese hamster ovary (CHO) cells are the expression system of choice for the industrial production of recombinant protein therapeutics due to their ability to grow to high densities in suspension culture, their robust nature, their track record in industry, their ability to produce human-like post-translational modifications (e.g. glycosylation) and their capability of producing high quality protein therapeutics (Omasa, Onitsuka, & Kim, 2010) (Wurm, 2004) (Walsh, 2014). As a result, there is significant scientific and commercial interest in research to significantly improve the efficiency of biotherapeutic production from CHO cells (Fischer, Handrick, & Otte, 2015). In recent years, the protein production capabilities of recombinant CHO cells have significantly improved (Bandaranayake & Almo, 2014), routinely producing high-quality protein in grams/L concentrations (Y.-M. Huang et al., 2010). Most of these improvements in productivity and product quality can be attributed to bioprocess optimisation approaches such as in media optimisation, feed development, bioreactor design, etc., and often these have been based on iterative empirical approaches which can be time-consuming and costly (Hacker et al., 2009)(Stolfa et al., 2017). In order to further improve the production capabilities and efficiency of CHO cells, an increased understanding of CHO cell biology is of the utmost importance. The next generation of improvements is expected to be made via genetic engineering of the host (CHO) cell itself to improve the efficiency of the production of biotherapeutic protein products (Fischer et al., 2015). Drug discovery pipelines of biopharmaceutical companies are producing ever-more complex molecules which are putting increasing pressure on cell line development groups to generate enough material to meet requirements for clinical trials. The recent sequencing of CHO genomes has opened up new possibilities for the application of systems biology to understand and

manipulate CHO cell biology (X. Xu et al., 2011b)(N. E. Lewis et al., 2013a)(Brinkrolf et al., 2013)(Kaas, Kristensen, Betenbaugh, & Andersen, 2015) and will enable a fundamental understanding of CHO cell physiology. This improved understanding of recombinant protein production will be instrumental in metabolic and cell engineering interventions to enhance CHO cell line (Datta, Linhardt, & Sharfstein, 2013) (Kildegaard, Baycin-Hizal, Lewis, & Betenbaugh, 2013).

The vast majority of 'omics based studies in Chinese hamster ovary (CHO) cells carried out to date have continued to miss out on the post-translational level of regulation. Phosphorylation events are crucial to the regulation of growth, cell cycle arrest, apoptosis, transcription, signal transduction, etc. (Yan & He, 2008), and hence are likely to be central to understanding and controlling bioprocess-relevant phenotypes. Post-translational modifications, either by proteolytic cleavage or by covalent modification of amino acids, can bring about changes in protein characteristics and functionality (Choudhary & Mann, 2010). Furthermore, genomic and transcriptomic studies are likely to reveal an incomplete picture of regulation if post-translational regulation of the final effector molecules, i.e., through a post-translational modification on the regulatory protein, is ignored (V. M. Tan, Cheng, & Drake, 2016). Hence, in order to gain a complete insight into the biology of CHO cells, a global differential study of post-translational modifications such as reversible phosphorylation of proteins is of great significance (Gutierrez & Lewis, 2015)(Henry, Power, et al., 2017b).

Post-translational modification of proteins by reversible phosphorylation on serine, threonine and tyrosine residues is a major switch mechanism of protein activity (Humphrey, James, & Mann, 2015). The importance of phosphorylation regulation is demonstrated in that up to 2% of the eukaryotic genome codes for proteins from the

kinase gene family (Caenepeel, Charydczak, Sudarsanam, Hunter, & Manning, 2004). The human genome contains approximately 568 protein kinases and 156 protein phosphatases to regulate phosphorylation events and play a crucial role in the control of biological processes such as proliferation and apoptosis (Ardito et al., 2017). Furthermore, up to 30% of all cellular proteins are estimated to be phosphorylated on at least one residue at any given time, though not all may be functionally relevant (P Cohen, 2000).

Mass spectrometry has now become the principal technique for global and targeted analysis of phosphoproteins and phosphorylation events (Vyse, Desmond, & Huang, 2017)(Savitski, Scholten, Sweetman, Mathieson, & Bantscheff, 2010). Despite the widespread phosphorylation of proteins, a functionally phosphorylated isoform of a particular regulatory protein is significantly lower than the total amount of the protein present in the cell; therefore the large-scale analysis of phosphoproteins/phosphopeptides is generally achieved by enrichment strategies such as immobilised metal ion ($Fe3^+$ or $Ga3^+$) affinity chromatography (IMAC) and metal oxide affinity chromatography (MOAC) using TiO_2 or ZrO_2 (Angel et al., 2012).

It is expected that proteomic and phosphoproteomic patterns will change during the different growth phases of a batch culture and these changes may give an indication of the cellular state regarding viability, growth, and productivity similar to mRNA and miRNA profiling studies of CHO cells grown in batch suspension culture (Bort et al., 2012). In this study, we have applied a well-established technique for phosphopeptide enrichment from CHO DP12 whole cell lysate samples. Immobilized metal affinity chromatography (IMAC; Fe_NTA) (Z. G. Wang, Lv, Bi, Zhang, & Ni, 2015) was used to enrich phosphopeptides from CHO DP12 cells prior to analysis by LC-MS/MS at different phases of cellular growth in serum-free suspension batch culture, i.e. during

the exponential phase, stationary phase and decline/death phase of culture. The data generated emphasises the dynamic nature of the changing proteome and phosphoproteome during the different growth phases over time in culture. The data also clearly demonstrates the importance of including multiple time points in experiments when comparing CHO cellular phenotypes or culture conditions (Bort et al., 2012).

2.2   Materials and methods

2.2.1   Assessment of cell number and viability.

IgG-producing CHO DP12 cells (purchased from the ATCC) were seeded at $3x10^5$ cells/mL in a 50 mL working volume containing CHO-S-SFM II serum free media (Thermo Fisher Scientific) in suspension batch culture at 37 ºC and 5 % CO_2 in a Kuhner Climico-Shaker ISF1-X orbital shaker at 170 rpm. A total of 12 flasks were seeded, and three replicate flasks were harvested for sample collection at each time point, i.e., at 48 hours, 72 hours, 96 hours and 120 hours after initial seeding, to capture early exponential, exponential, stationary and death phase of culture based on the growth profiling experiments. Cellular growth and culture viability were assayed using a Guava Benchtop Cytometer after staining with Viacount (Millipore). All cells were tested for *Mycoplasma* using the Hoeschst indirect culture test carried out in-house and were found to be free of contamination.

2.2.2   Cell lysate preparation

Cells were pelleted by centrifugation, washed with cold PBS and were stored at -80ºC until required. Cells were resuspended in 8 M urea, 50 mM Tris, 75 mM NaCl (pH 8.2) buffer supplemented with 1X Halt protease inhibitors (Thermo Fisher Scientific), 1X Halt phosphatase inhibitors (Thermo Fisher Scientific), and 0.1% ProteaseMAX surfactant detergent (Promega). Cells were sonicated at a medium setting for 30

seconds for a total of three pulses or until a homogeneous, free-flowing mixture of cell lysate was observed with at least a 5 minute rest between each sonication pulse. The cell lysate was centrifuged at 14,000 x*g* for 10 minutes at 4 ºC and the supernatant was transferred to a clean microcentrifuge tube. Protein quantitation was carried out using the Quick Start Bradford Protein Assay (BioRad). The remaining supernatant was stored at -80 ºC.

2.2.3   In-Solution Protein Digestion and Phosphopeptide Enrichment

2 mg of each cell lysate protein preparation was transferred to a new centrifuge tube, and all samples were equalised to the same volume using the same resuspension buffer as described above. A fresh stock of 0.5 M dithiothreitol (DTT) was prepared, and an appropriate volume of DTT was added to achieve a final concentration of 5 mM. Samples were reduced for 25 minutes at 56 ºC. Before alkylation, samples were allowed to cool to room temperature and an appropriate volume of freshly prepared 0.5 M iodoacetamide was added to a final concentration of 14 mM. Samples were alkylated for 30 minutes at room temperature in the dark. Protein samples were diluted at a ratio of 1:5 in 25 mM Tris-HCl, pH 8.2, to reduce the concentration of urea to less than 1.6 M. Samples were digested with sequence grade modified trypsin (Thermo Fisher Scientific) at a ratio of 1:100 enzyme:substrate for 4 hours at 37 °C. Fresh trypsin at the same ratio was then added to each sample for an overnight digestion at 37 °C. After overnight digestion, TFA was added to a final concentration of 0.4% to inactivate the trypsin. Digested samples were desalted and concentrated using a Sep-Pak C18 Vac cartridge (Waters) with negative pressure. The C18 cartridge was washed and conditioned by adding 9 mL of acetonitrile (ACN) followed by 3 mL of 50% ACN and 0.5% acetic acid. The cartridge was then equilibrated with 9 mL of 0.1% TFA, and the samples were loaded in 0.4% TFA. Loaded samples were desalted with 9 mL

0.1% TFA. TFA was removed with 1 mL 0.5% acetic acid. Desalted peptides were eluted with 6 mL of 50% ACN, 0.5% acetic acid. The eluates from each sample were snap frozen with liquid nitrogen, lyophilized by freeze drying to dry powder and stored at -20 ºC. 10 µg of digested protein was set aside for non-enriched sample analysis. 2 mg of tryptically digested protein samples were phosphopeptide-enriched using Fe-NTA (IMAC) spin columns (Pierce, Thermo Fisher Scientific) as per manufacturer's instructions.

Non-enriched peptide and phosphopeptide sample concentrations were determined using a Nanodrop One (Laptech International, UK).

2.2.4 LC-MS/MS Analysis

LC-MS/MS separations were performed using an UltiMate 3000 nanoRSLC system (Thermo Scientific) coupled in-line with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific).

For non-enriched samples, 1.3 µL of cell lysate samples (equivalent to 1 µg of digested peptides) were loaded onto the trapping column (PepMap100, C18, 300 µm × 5 mm) (Thermo Scientific) for 3 minutes at a flow rate of 25 µL/min with 2% (v/v) ACN, 0.1% (v/v) TFA. Peptides were resolved on an analytical column (Easy-Spray C18 75 µm × 250 mm, 2 µm bead diameter column) (Thermo Scientific) using a gradient of 98% A (0.1% (v/v) formic acid (FA)): 2% B (80% (v/v) ACN, 0.08% (v/v) FA) to 35% B over 120 min at a flow rate of 300 nL/min.

MS(/MS) data were acquired on the Orbitrap Fusion Tribrid as follows for the non-enriched whole cell lysate samples: all MS1 spectra were acquired over m/z 400−1500 in the Orbitrap (120 K resolution at 200 m/z), automatic gain control (AGC) was set to accumulate 4×10^5 ions with a maximum injection time of 100 ms. Data-dependent tandem MS analysis was performed using a top-speed approach (cycle time of 3 s),

and the normalized collision energy was optimized at 35% for Collision Induced Dissociation (CID). MS2 spectra were acquired with a fixed first m/z of 100. The intensity threshold for fragmentation was set to 5000 and included charge states 2+ to 7+. A dynamic exclusion of 50 s was applied with a mass tolerance of 10 ppm.

For the phosphopeptide-enriched samples, 2.5 μL of sample (equivalent to 1 μg of digested phosphopeptides) were loaded onto the trapping column (PepMap100, C18, 300 μm × 5 mm), using partial loop injection, for 3 min at a flow rate of 25 μL/min with 2% (v/ v) ACN, 0.1% (v/v) TFA and then resolved on an analytical column (Easy-Spray C18 75 μm × 250 mm, 2 μm bead diameter column) using a gradient of 98% A (0.1% (v/v) formic acid (FA)): 2% B (80% (v/v) ACN, 0.0.08% (v/v) FA) to 35% B over 120 min at a flow rate of 300 nL/min. For the phosphopeptide-enriched samples MS(/MS) data were acquired on the Orbitrap Fusion Tribrid as follows: all MS1 spectra were acquired over m/z 400−1400 in the Orbitrap (120 K resolution at 200 m/z), automatic gain control (AGC) was set to accumulate 4×10^5 ions, with a maximum injection time of 50 ms. Data-dependent tandem MS analysis was performed using a top-speed approach (cycle time of 3 s) and the normalized collision energy CID was optimized at 35%. MS2 spectra were acquired with a fixed first m/z of 100. The intensity threshold for fragmentation was set to 10,000 and included charge states 2+ to 6+. A dynamic exclusion of 50 s was applied with a mass tolerance of 10 ppm. Multistage activation was enabled for neutral-loss triggered fragmentation for all precursor ions exhibiting neutral loss of mass 97.9763 or 80 Da with a mass tolerance of 0.5 m/z, where the neutral loss ion was one of the top 10 most intense MS2 ions. AGC was set to 20,000 with a maximum injection time set at 90 ms.

2.2.5 Quantitative Label-free LC-MS/MS Analysis

Differential quantitative label-free LC-MS/MS data analysis was carried by interrogating the resultant LC-MS/MS files using the software package Progenesis QI for Proteomics (NonLinear Dynamics, Waters) essentially as previously described (Henry, Power, et al., 2017a). Raw MS data files from each of the 12 samples were imported into the Progenesis QI software interface, and automatic reference run alignment was carried out to account for variability between runs. Upon alignment of all runs, identified features were filtered based on ANOVA p-value < 0.05 between experimental groups. For proteomic and phosphoproteomic analysis, a mascot generic file (mgf) was generated from all exported MS/MS spectra and analysed using Proteome Discoverer v.2.1 (Thermo Fisher Scientific) in conjunction with SEQUEST. Peak lists were searched against an NCBI-Chinese hamster ovary (*Cricetulus griseus*) protein database (fasta file downloaded 29th November 2017 containing 24,906 sequences). Parameters were set as follows: MS1 tolerance of 10 ppm; MS2 mass tolerance of 0.6 Da for ion trap detection; enzyme specificity was set as trypsin with two missed cleavages allowed; carbamidomethylation of cysteine was set as a fixed modification; and phosphorylation of serine, threonine, and tyrosine (for phosphopeptide analysis) and oxidation of methionine were set as variable modifications. For phosphosite identification, the Phospho*RS* algorithm (Taus et al., 2011) was run through ProteomeDiscover 2.1 using diagnostic fragment ions and analyzer-specific fragment ion tolerances, as described above. Data were filtered to a 1% false discovery rate (FDR) on PSMs using automatic decoy searching in SEQUEST and by applying a phosphosite probability score of 75% or greater for S, T or Y amino acids in Phospho*RS* (Taus et al., 2011). Once peptides and phosphopeptides were successfully identified, a statistical criteria of ANOVA p-value

≤ 0.05 and a relative minimum fold change of ±1.5 between the experimental groups were set. Only peptides and phosphopeptides which passed these criteria were then deemed to be differentially expressed between the relevant experimental groups being analysed.

2.2.6 Gene Ontology Analysis

Identified differentially expressed proteins and phosphoproteins were assigned to mouse or human official Gene Symbol identifiers. These gene symbols were then imported into DAVID (https://david.ncifcrf.gov), ToppGene Suit (https://toppgene.cchmc.org) and Enrichr (http://amp.pharm.mssm.edu/Enrichr/) online gene ontology platforms for functional pathway analysis. To refine the number of pathways identified, an adjusted p-value (Benjamini-Hochberg) of ≤0.05 between experimental groups was used. For each phosphoprotein of interest, the phosphosite number was assigned from the corresponding human protein sequence match. The functional information of such phosphosites was obtained from phosphosite.org (www.phosphosite.org), an online resource for providing information and tools for the study of protein post-translational modifications. For phosphopeptides with no complementary human sequence match, the phosphosite number was assigned through CHO-K1 protein database (CHO-K1[ATCC]_RefSeq_2014) (www.chogenome.org).

2.3 Results

2.3.1 Proteomic and phosphoproteomic analysis of the growth of CHO cells in batch culture

In this comparative proteomic and phosphoproteomic study of the IgG producing CHO DP12 cells, we aimed to identify differentially expressed proteins and phosphoproteins as the cells transition through the exponential, stationary and decline

phases of batch growth in suspension-cultured CHO cells. The cells were harvested on Day 2 and then after every 24 hours until Day 5, and are referred to as conditions Day 2, Day 3, Day 4 and Day 5. These time points were selected based on our understanding of the behaviour of CHO DP12 cells in serum-free media in suspension batch culture. During this short culture duration selected time points capture the culture transition from early to late exponential phase (Day 2 to Day 3), through to the stationary phase (Day 3 to Day 4) and ultimately to the decline phase of the cell culture (Day 4 to Day 5). Figure 2.1A shows the growth and culture viability of the CHO DP12 cells used for this study.

For each condition or day, 2 mg of protein was tryptically digested. 10 μg from each digested sample was analysed as the whole cell lysate non-enriched sample and the remainder was used for phosphopeptide enrichment. Each sample was analysed by LC-MS/MS. Raw MS files generated from the Orbitrap Fusion Tribrid mass spectrometer were interrogated using Progenesis QI for Proteomics in order to identify statistically significant differentially expressed proteins and phosphoproteins over the culture duration. From this LC-MS data, we identified approximately 4,900 proteins in total from the 2 hour LC-MS runs from the non-enriched whole cell lysate samples and approximately 7,200 phosphopeptides in total from the phosphopeptide-enriched fractions.

To best understand the progressive changes in the proteome and the phosphoproteome of the CHO DP12 cells over time in batch suspension culture, LC-MS/MS data files from the three replicate samples from two consecutive day samples were compared with each other to obtain an account of differentially expressed peptides and phosphopeptides that may have changed in abundance during the transition from each day of culture to the next, i.e., Day 2 vs. Day 3, Day 3 vs. Day 4, and Day 4 vs. Day

5. Significantly increased or decreased expression of phosphopeptides between samples was determined by using a fold-change cut-off of ±1.5 fold and a p-value ≤0.05 between experimental groups (Henry, Power, et al., 2017b).   The phosphorylation site in the peptide was localised through phospho*RS* 3.1 in ProteomeDiscoverer 2.1 after applying a phosphosite probability cut-off score of ≥75% (Taus et al., 2011). Peptides from the non-enriched whole cell lysate samples with fold changes ±1.5 and a p-value ≤0.05 between experimental groups were also deemed to be differentially expressed. Figure 2.2A provides an overview of the numbers of differentially expressed peptides and phosphopeptides and their corresponding proteins over time in culture. Unsupervised Euclidian clustering analysis of the differentially expressed phosphopeptides and non-enriched peptides show a clear clustering of each sample set into the different days of the batch suspension culture, as shown in Figure 2.1 B, i.e. Day 2, Day 3, Day 4, and Day 5 of the cell culture. There is some variation between the samples from each time point; however, the Unsupervised Euclidian clustering of the phosphoproteomic and total proteomic data clustered the three biological replicates from each time point into distinct sample groups. This relationship can be confirmed from a Principal Component Analysis (PCA) plot of the phosphoproteomic and proteomic data, as shown in Supplementary Figure 1.
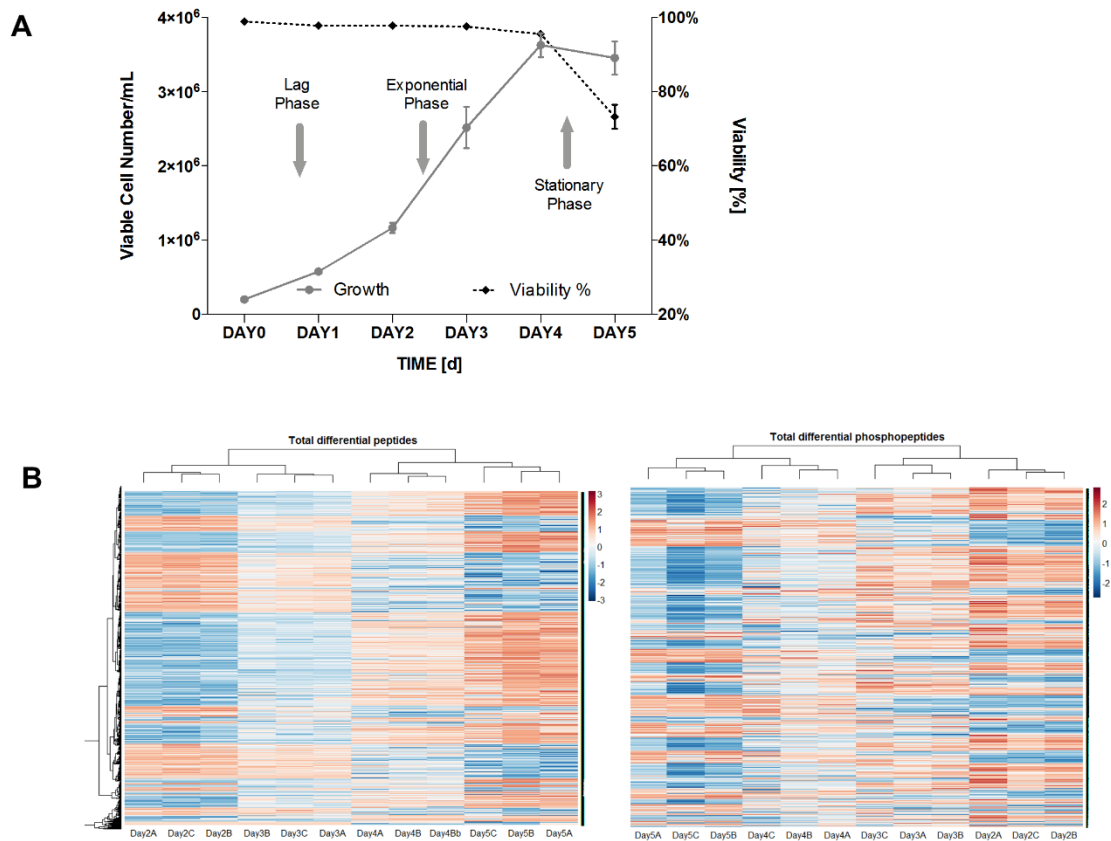
Figure 2.1 Growth curve of CHO DP12 cell line and heat maps of differentially expressed peptides and phosphopeptides over time in culture. (A) Cell number and culture viability (%) over time in serum-free suspension batch culture. Samples for proteomic and phosphoproteomic analysis were captured on day 2 (early exponential), day 3 (exponential), day 4 (stationary) and day 5 (early death) phases of growth. (B) Unsupervised Euclidian clustering shows that the expression of the differentially expressed non-enriched peptides (left) and phosphopeptides (right) identified from the experimental samples separate into four distinct sample groups i.e. Day 2, Day 3, Day 4, and Day 5 of the cell culture. Three replicate samples (labelled A-C) are shown for each time point.

In this analysis, 1867 phosphopeptides were found to be differentially expressed between Day 2 and Day 3 of culture as the cells transition from early through to late exponential phase of growth, with 923 phosphopeptides showing increased expression and 944 phosphopeptides showing a decrease in expression at Day 3 of culture compared to Day 2. In the comparison between Day 3 and Day 4, 1096 phosphopeptides were differentially expressed, with 370 phosphopeptides showing increased expression at Day 4 and 726 phosphopeptides showing decreased expression at Day 4 of culture as the cells transition from the exponential to the stationary phase.

Similarly, differential analysis of Day 4 vs Day 5 of the culture identified 1521 differentially expressed phosphopeptides with 407 phosphopeptides showing increased expression at Day 5 and 1114 phosphopeptides showing decreased expression at Day 5 as the cells transition into the decline/death phase of growth. A summary of the total number of differentially expressed phosphopeptides and peptides and their corresponding proteins is shown in Figure 2.2A. Figure 2.2B shows the overlap between the phosphoproteomic and proteomic differentially expressed proteins for each comparative analysis. The full lists of differentially expressed phosphopeptides and their corresponding phosphoproteins are shown in Supplementary Table C2-1. The full lists of differentially expressed non-enriched peptides and the corresponding non-enriched proteins from the whole cell lysate (WCL) samples are shown in Supplementary Table C2-2.

A number of phosphopeptides were differentially expressed between multiple comparisons; for example, 379 phosphopeptides were common between the Day 2 vs. Day 3 and Day 3 vs. Day 4 comparisons, and 428 phosphopeptides were commonly identified between and Day 3 vs. Day 4 and Day 4 vs. Day 5 comparative analyses. Interestingly, 111 phosphopeptides corresponding to 97 phosphoproteins were common between all comparison sets. Such phosphopeptides represent the dynamic nature of the phosphoproteome and give an account of potential acute changes in phosphorylation in response to cell growth. A change diagram of these constitutively differentially phosphopeptides is shown in Figure 2.3, showing the phosphopeptides that generally increase in expression over time in culture as Cluster 1 and phosphopeptides that generally decrease over time in culture as Cluster 2. A list of phosphopeptides that were observed to be differentially expressed at all phases of

growth is included in Supplementary Table C2-3 and includes reference to whether

the phosphopeptides belong to Cluster 1 or Cluster 2.

A Comparison	Phosphopeptides	Phosphoproteins	Non-enriched peptides	Non-enriched proteins
DAY 2 vs DAY3 (Early to late exponential)	1867 923 ↑ Day3 944 ↓ Day3	835 440 ↑ Day3 395 ↓ Day3	1332 418 ↑ Day3 914 ↓ Day3	311 219 ↑ Day3 92 ↓ Day3
DAY3 vs DAY4 (Exponential to Stationary)	1096 370 ↑ Day4 726 ↓ Day4	585 213 ↑ Day4 372 ↓ Day4	1873 1075 ↑ Day4 798 ↓ Day4	432 214 ↑ Day4 218 ↓ Day4
DAY4 vs DAY5 (Stationary to Decline)	1521 407 ↑ Day5 1114 ↓ Day5	754 214 ↑ Day5 540 ↓ Day5	1465 934 ↑ Day5 531 ↓ Day5	367 209 ↑ Day5 158 ↓ Day5

B

Phosphoproteins — Total Proteins — 793 | 42 | 269 — **DAY2 vs DAY3**

Phosphoproteins — Total Proteins — 541 | 44 | 388 — **DAY3 vs DAY4**

Phosphoproteins — Total Proteins — 693 | 61 | 306 — **DAY4 vs DAY5**

Figure 2.2 Overview of the total number of differentially expressed phosphopeptides, and peptides for each comparative analysis. (A) Summary of differentially expressed peptides, phosphopeptides and their corresponding proteins, and their direction of fold change. (B) Venn diagrams showing the overlap between the proteomic and phosphoproteomic protein identifications between the three comparative analyses. Peptides and phosphopeptides and their corresponding proteins were deemed differentially expressed if the fold change between two conditions was ±1.5 with an ANOVA p-value ≤ 0.05 between experimental groups.
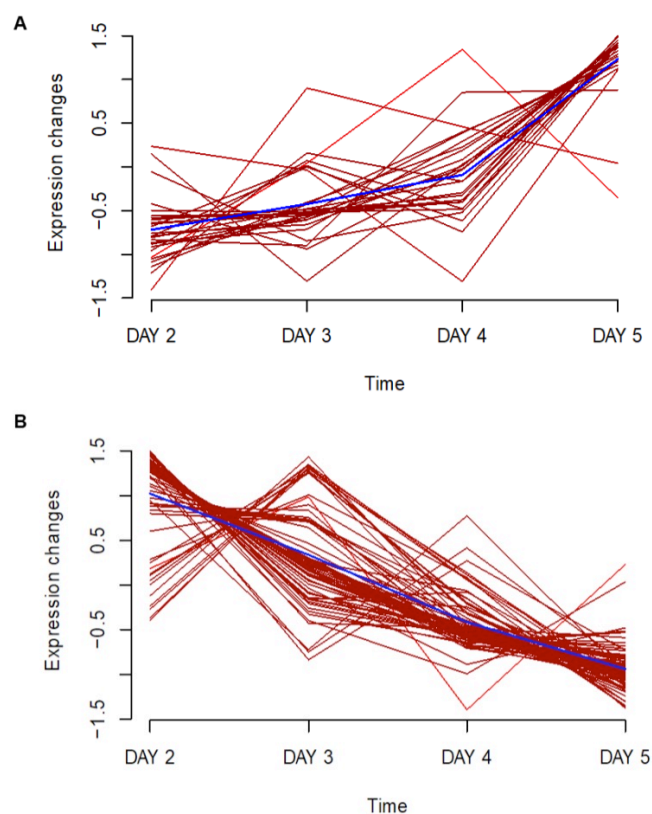
Figure 2.3 A change diagram of phosphopeptides shown to be differentially expressed at all growth phases. Unsupervised soft clustering of phosphopeptides revealed two different cluster structures, (A) Cluster 1 and (B) Cluster 2. Log10 transformed phosphopeptide abundance values on the y-axis are plotted against the sample collection time points on the x-axis.

2.4 Overlap of proteomic and phosphoproteomic datasets

A significant limitation in phosphoproteomic studies is the ambiguity of total peptide stoichiometry concerning its phosphorylated isoform (Solari, Dell'Aica, Sickmann, & Zahedi, 2015). As a result one of the major challenges in phosphoproteomics is that abundance differences of a phosphopeptide could arise from (1) differential phosphorylation or (2) the differential overall expression of the total phosphoprotein. A comparison between the relative abundance pattern of a phosphopeptide and the non-phosphorylated version of the same peptide, if identified, can provide evidence for differential phosphorylation/dephosphorylation rather than just a change in abundance of a phosphopeptide and hence the phosphoprotein itself. With rapidly improving capabilities of high resolution and highly sensitive mass spectrometers such

as the Orbitrap Fusion™ Tribrid™ mass spectrometer used in this study, a good overlap between phosphopeptide identification and total proteome identification is expected. Supplementary Table C2-4 shows the list of all differentially expressed phosphopeptides (i.e. 532 in total) for which the corresponding peptide was identified in the non-enriched whole cell lysate, but not differentially expressed, suggesting altered phosphorylation of these proteins over the growth phases in batch culture. From this list of 532 phosphopeptides, three examples are shown in Figure 2.4 to demonstrate differential expression of a phosphopeptide with no change in the corresponding peptide from the total protein. For example the phosphopeptide S*GEGEVSGLMR, with identified phosphosite corresponding to Ser473 of human Transcription intermediary factor 1-beta (TRIM28), was found to be differentially expressed at all phases of growth, whereas the non-phosphorylated version of the peptide did not change over time in culture (see Figure 2.4A). The phosphopeptide abundance varies significantly across the growth curve with relative abundance firstly decreasing 1.9 fold from day 2 to day 3 and then gradually rising from day 3 until the end of the culture duration. However, no significant fold changes were observed from the analysis of the non-phosphorylated version of the peptide over time in culture. It has previously been shown that depletion of TRIM28 using shRNAi leads to increased cellular proliferation (L. Chen et al., 2012) and phosphorylation of TRIM28 on Ser473 by MAP kinase-activated protein kinase 2 (MK2) has been shown to reduce the activity of TRIM28 (King, 2013). These previous studies tie in with the phosphoproteomic data shown here with the initial decrease in phosphorylation during the exponential phase of growth (day 2 to day 3) and then an increase in phosphorylation as the growth rate decreases and eventually the culture entering the decline phase by day 5. Similarly, the phosphopeptide QAGPS*PEAELR of protein

Pleckstrin-like domain-containing family G member 6 (PLEKHG6), which is a guanine nucleotide exchange factor that can activate RhoG and RhoA and plays a role in cytokinesis (D. Wu, Asiedu, Adelstein, & Wei, 2006)(Samson, Welch, Monaghan-Benson, Hahn, & Burridge, 2010), was identified to be differentially expressed at all phases of growth, initially showing a decrease in expression by 1.8 fold between day 2 and day 3, and then increasing in expression 1.8 fold from day 3 to day 4 and continuing to increase (1.7 fold) from day 4 to day 5. However, the abundance of the non-phosphorylated peptide QAGPSPEAELR identified from the non-enriched whole cell lysate did not change significantly over the culture duration (see Figure 2.4B). Similarly, phosphopeptide VSHYIINSSGPRPPVPPSPAQPPPGVSPS*R (Ser85) identified from Proto-oncogene c-CRK (also known as Adaptor molecule CRK), which belongs to a family of adaptor proteins that act as major convergence points of tyrosine kinase signaling pathways (Braiman & Isakov, 2015), shows a 2 fold decrease from day 2 to day 3, a 1.8 fold decrease from day 3 to day 4, and a 1.9 fold decrease from day 4 to day 5 of the CHO cell culture, whereas the non-phosphorylated version of the peptide shows no significant change in abundance over time in culture (see Figure 2.4C).
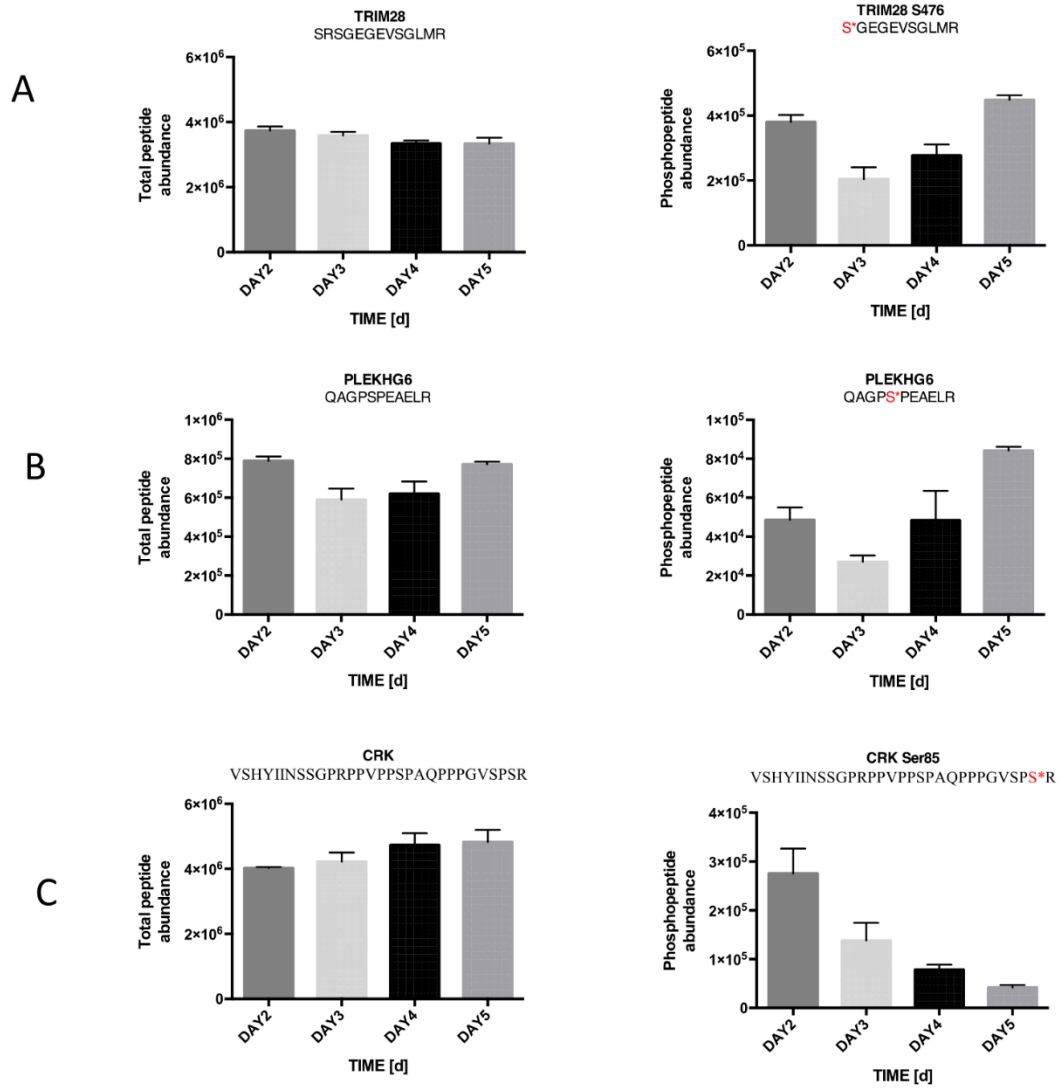
Figure 2.4 Altered phosphorylation of phosphoproteins during batch suspension culture. Expression patterns of peptides and phosphopeptides from (A) TRIM28, (B) PLEKHG6, and (C) CRK over time in culture. Left panel shows the change in expression of the non-phosphorylated peptide identified from the non-enriched whole cell lysate over time in culture. Right panel show changes in expression of the phosphorylated version of the same peptide over time in culture. X-axis, days of cell culture; Y-axis, peptide/phosphopeptide abundance from MS data.

2.5 Functional classification of significantly differentially expressed proteins and phosphoproteins

All differentially expressed proteins were assigned human official gene identifiers for functional classification. Functional annotation of these genes was performed using DAVID (https://david.ncifcrf.gov) and Toppgene suite for functional analysis (https://toppgene.cchmc.org). For refined identification of statistically significant

67

enrichment terms, an adjusted p-value (Benjamini Hochberg) of ≤0.05 was used. The most significant of the GO: Biological Process terms identified from the phosphoproteomic study were related to cell cycle, RNA processing, regulation of organelle organization, cytoskeleton organization, chromatin organization, histone modifications, etc. The molecular function of these proteins indicated that the differentially expressed phosphoproteins were of importance in, for example, transcription factor and transcription cofactor activity, transcription factor binding, GTPase regulator activity and kinase binding. (See Supplementary Table C2-5 for the top 20 most-enriched GO: Biological Processes and GO: Molecular Functions from the phosphoproteomic and the non-enriched whole cell lysate proteomic analyses). We also identified the differential expression of 94 kinases along the growth curve, as shown in Supplementary Table C2-6, which is a remarkable number given that the human kinome contains close to 518 kinases (Duong-Ly & Peterson, 2013).

From KEGG pathway analysis we observed the enrichment of a number of pathways during the various phases of growth in batch culture from analysis of the lists of differentially expressed phosphoproteins. For example, there was a significant enrichment of growth-related signalling pathways such as ErBb signalling, insulin signalling and cancer-related pathways. Table 2.1 lists the KEGG pathways that were identified to be active at various phases of cell growth. Two of the pathways that were found to be highly enriched in the exponential and stationary phases were the mTOR and autophagy pathways. The mTOR pathway has been manipulated previously in CHO cells to improve production efficiencies (Dadehbeigi & Dickson, 2015)(Josse et al., 2016)(Dreesen & Fussenegger, 2011), and recently autophagy in CHO cells have gained attention as an opportunity for controlled cell death in bioprocess (Baek, Kim, Kim, Lee, & Lee, 2016). In our study, we identified 18 phosphoproteins from the

mTOR pathway, including the components of both mTORC1 and mTORC2 complexes. We also observed the differential expression of 17 autophagy-related phosphoproteins, predominantly at the exponential phase of cell culture; Supplementary Table C2-7 provides the complete list of differentially expressed phosphopeptides and associated phosphosites identified from the mTOR and autophagy pathways. Supplementary Figure 2 shows KEGG pathways highlighting the differentially expressed phosphoproteins from the mTOR and autophagy pathways identified in this study. Three of the phosphosites identified from WIPI2 and ATG2A from the autophagy pathway are potentially CHO-specific as no equivalent phosphosites were found in human sequences.

KEGG pathway analysis of the non-enriched whole cell lysate differentially expressed proteins showed a different set of enriched pathways and included an enrichment of proteins related to metabolism such as the Citrate (TCA) cycle, carbon metabolism, amino acid synthesis and degradation (see Table 1.1). This again demonstrates the extra depth of information achieved when combining both the proteomic and phosphoproteomic datasets.

Table 2.1 Functional pathway enrichment of proteins and phosphoproteins differentially expressed during the serum-free suspension batch culture.

	KEGG Pathway	pValue	BH Adjusted	Count
Phosphoproteome				
Day 2 vs Day 3	Spliceosome	1.44E-10	3.90E-08	28
	ErbB signaling pathway	6.46E-06	8.79E-04	16
	RNA transport	2.11E-05	1.24E-03	23
	Thyroid hormone signaling pathway	2.41E-05	1.24E-03	18
	Insulin signaling pathway	2.43E-05	1.24E-03	20
	Autophagy - animal	2.74E-05	1.24E-03	19
	Ribosome biogenesis in eukaryotes	9.48E-05	3.68E-03	16
	Non-small cell lung cancer	1.50E-04	5.09E-03	11
	mTOR signaling pathway	2.63E-04	7.94E-03	19
Day 3 vs Day 4	AMPK signaling pathway	1.72E-06	4.04E-04	17
	Autophagy - animal	3.78E-06	4.45E-04	17
	Cell cycle	1.06E-05	8.30E-04	16
	mTOR signaling pathway	3.47E-05	1.92E-03	17
	Insulin signaling pathway	4.09E-05	1.92E-03	16
	Spliceosome	1.06E-04	4.16E-03	15
	Progesterone-mediated oocyte maturation	1.82E-04	5.93E-03	12
	Endocytosis	2.10E-04	5.93E-03	22
Day 4 vs Day 5	RNA transport	2.71E-06	6.84E-04	22
	Adherens junction	4.02E-05	5.06E-03	12
	AMPK signaling pathway	1.59E-04	1.34E-02	15
	Tight junction	7.89E-04	4.97E-02	17
Proteome				
Day 2 vs Day 3	Spliceosome	2.45E-29	8.57E-27	64
	RNA transport	7.23E-27	1.27E-24	70
	Metabolic pathways	2.16E-20	2.53E-18	228
	Carbon metabolism	2.03E-19	1.78E-17	48
	Protein processing in endoplasmic reticulum	4.81E-12	3.37E-10	48
	Biosynthesis of amino acids	6E-12	3.5E-10	30
	Proteasome	3.37E-11	1.69E-09	22
	Ribosome	5.79E-11	2.53E-09	44

	Valine, leucine and isoleucine degradation	1.67E-10	6.5E-09	22
	Citrate cycle (TCA cycle)	2.89E-10	1.01E-08	17
Day 3 vs Day 4	Carbon metabolism	4.66E-26	1.64E-23	56
	Metabolic pathways	1.29E-24	2.27E-22	244
	Spliceosome	2.14E-23	2.51E-21	58
	Citrate cycle (TCA cycle)	4.45E-19	3.92E-17	24
	RNA transport	1.11E-18	7.82E-17	60
	Citrate cycle (TCA cycle, Krebs cycle)	1.52E-16	8.92E-15	19
	Protein processing in endoplasmic reticulum	4.76E-14	2.4E-12	52
	Proteasome	5.76E-14	2.54E-12	25
	Valine, leucine and isoleucine degradation	4.15E-13	1.63E-11	25
	Biosynthesis of amino acids	1.8E-12	6.37E-11	31
	Phagosome	5.23E-07	9.23E-06	37
Day 4 vs Day 5	Carbon metabolism	4.31E-25	1.48E-22	48
	Metabolic pathways	1.6E-19	2.75E-17	181
	Citrate cycle (TCA cycle)	3.06E-19	3.51E-17	22
	Citrate cycle (TCA cycle, Krebs cycle)	8.73E-16	7.51E-14	17
	RNA transport	4.11E-15	2.83E-13	46
	Valine, leucine and isoleucine degradation	3.97E-13	2.28E-11	22
	Spliceosome	8.65E-13	4.25E-11	37
	Citrate cycle, second carbon oxidation,	1.86E-12	8.01E-11	12
	Protein processing in endoplasmic reticulum	2.79E-12	1.07E-10	41
	Ribosome biogenesis in eukaryotes	3.27E-10	1.13E-08	29
	Proteasome	2.38E-06	3.90E-05	14

2.6 Discussion

The data in this study shows significant changes in the proteome and phosphoproteome during growth of CHO cells in suspension batch culture. The entire study identified 3,777 differentially expressed CHO phosphopeptides corresponding to 1,415 differentially expressed phosphoproteins and 834 differentially expressed proteins over the culture duration. We also identified 94 differentially expressed kinases in the study. The regulation of phosphorylation by kinases and phosphatases allows cells to respond to internal and external signaling events on the fly (Humphrey et al., 2015).

From the gene ontology analysis of the phosphoproteomic datasets we identified a significant enrichment of proteins related to transcription factor activity; this enrichment was not observed in the whole cell lysate proteomic data. Phosphorylation is known to target gene transcription by affecting the stability, location and structure of transcription factors (Vogel, Gafken, Leid, & Filtz, 2014). Interestingly gene ontology analysis of the whole cell lysate samples showed an enrichment of molecular functions primarily associated with metabolic pathways such as the TCA cycle, carbon metabolism, amino acid metabolism, etc. as the culture progresses over time. An improved knowledge of such metabolic pathways in CHO cells which can lead to the production and accumulation of growth-inhibiting and toxic metabolites will lead to improved growth and productivity of CHO cells (Pereira, Kildegaard, & Andersen, 2018a).

One of the pathways that were found to be highly enriched in this study was the mTOR pathway. This pathway is of high interest in CHO bioprocessing where a number of studies have involved manipulation of the mTOR pathway to improve CHO cell growth and productivity(Dadehbeigi & Dickson, 2015)(Josse et al., 2016)(Dreesen &

Fussenegger, 2011). In this study, 18 phosphoproteins from the mTOR pathway were identified, including components of both mTORC1 and mTORC2 complexes as outlined in Supplementary Table C2-7. We identified the peptide TDS*YSAGQSVEILDGVELGEPAHK with a phosphorylated serine corresponding to Ser2248 (human) of the mTOR protein. The phosphorylation of Ser2248 on mTOR by p70S6K has been shown to be responsible for nuclear translocation on the mTORC1 complex (Leal et al., 2013). In this study we observed that total mTOR protein levels followed an upward trend from day 2 until day 5 (fold change >7) of culture, however the expression of phosphopeptide TDS*YSAGQSVEILDGVELGEPAHK (Ser2248) from mTOR shows only a 1.5 fold increase from Day 2 to Day 3 of the cell culture and thereafter following a downward trend up until day 5.

We also identified differentially expressed phosphopeptides related to two crucial downstream substrates 4E-BP1 and ULK1 of the mTORC1 complex. 4E-BP1 is a well-known substrate of the mTORC1 complex (Hara et al., 1997) and the phosphorylation status of 4E-BP1 has been used as a marker of mTOR activity (X. Qin, Jiang, & Zhang, 2016). We identified three functionally relevant (human) phosphosites i.e. Thr41, Ser65 and Thr70 on protein 4E-BP1. Activated mTORC1 hyperphosphorylates 4E-BP1 on multiple sites including Ser65 and Thr70 which lead to a release of eIF4E from 4E-BP1/eIF4E complex and initiation of cap-dependent translation (Gingras et al., 2001). We observed a 2 fold upregulation of a phosphopeptide (Thr70) from day 2 to day 3 of culture and a 4.5 fold decrease in the expression of a phosphopeptide containing the equivalent of Ser65 from day 3 to day 4 of culture. While the potential increase in phosphorylation of Thr70 indicates increased mTOR activity and a resultant increase in cap-dependent translation, the

dephosphorylation of Ser65 from day 3 to day 4 indicates a potential decrease in mTOR activity. These observations potentially show that the highest mTOR activity is registered on day 3 of the batch culture, which correlates with the phosphopeptide containing Ser2248 on mTOR being the most abundant at day 3. Increased phosphorylation of Ser2248 has been shown to be observed upon treatment of adult cardiac muscle cells with mTOR agonists such as endothelin-1 (ET-1), phenylephrine (PE), TPA, or insulin (Moschella, Rao, McDermott, & Kuppuswamy, 2007), and mTOR activity has been shown to be negatively regulated by the dephosphorylation of Ser2248 triggered by overexpression in cultured adipocytes (Guntur, Guilherme, Xue, Chawla, & Czech, 2010). The increased activity of mTOR at day 3 of cell culture is also indicated by the higher relative abundance of the phosphopeptide ILDTSSLTQSAPAS*PTNK (Ser863) at day 3 as Ser863 on RAPTOR has been shown to be required for positive regulation of mTORC1 activity (Foster et al., 2010). RAPTOR functions as a scaffold for mTORC1 complex substrates (Foster et al., 2010). Like RAPTOR, PRAS40 and other members of the mTORC1 complex can also regulate mTOR activity (Yuan & Guan, 2016). The phosphorylation of Thr246 of PRAS40 by PI3K has been shown to induce dissociation of PRAS40 from mTOR thus inhibiting mTOR activity (Wiza, Nascimento, & Ouwens, 2012). We found that the phosphopeptide LNT*SDFQK containing phosphosite Thr246 showed a 5.3 fold decrease from day 2 to day 5 of culture. Similar changes in abundance patterns were observed for phosphopeptide SSDEENGPPS*SPDLDR (Ser211) and SSDEENGPPSS*PDLDR (Ser212) of PRAS40. As a central pathway for cell proliferation and protein synthesis, the mTOR pathway has been manipulated in CHO cells to understand and improve bioprocess attributes. The phosphoproteomic data shows a much deeper insight into the complexity of the mTOR pathway and the

potential importance of post-translational regulation by phosphorylation of the mTOR pathway in recombinant CHO cells that would not be observed in proteomic studies. Pathway analysis of phosphoproteomic data also revealed a significant enrichment of autophagy-related proteins with differential expression between day 2 and day 4 of the batch suspension culture. Recently autophagy has gained importance in bioprocess as an opportunity to genetically and chemically control cell survival and productivity (Y. K. Han, Ha, Lee, Lee, & Lee, 2011)(Y. J. Kim, Baek, Lee, & Lee, 2013)(Nasseri et al., 2014)(Baek et al., 2016). It is therefore important to have a detailed understanding of the autophagic pathway in order to select relevant genes and chemical targets to manipulate autophagy in CHO cells to enhance the production of biotherapeutics (Y. J. Kim et al., 2013). To increase culture longevity and increase protein production, numerous attempts have been made to delay apoptosis of CHO cells (Druz, Son, Betenbaugh, & Shiloach, 2013)(J. G. L. Tan et al., 2015), however little is understood regarding the regulation of autophagy in CHO cells in bioprocess-relevant conditions. In our study we have observed the differential expression of 17 autophagy-related phosphoproteins, predominantly at the exponential phase of cell culture. For example, the phosphorylation status of the ULK1–ATG13–FIP200–ATG101 is crucial for autophagy initiation (Ganley et al., 2009). In this study we identified differentially expressed phosphopeptides from ULK1 and ATG13. Studies have shown ULK1 and ATG13 as direct substrates of the mTORC1 complex and the inhibition of mTORC1 by rapamycin of starvation conditions leads to dephosphorylation of these proteins (Puente, Hendrickson, & Jiang, 2016)(Kamada et al., 2010)(J. Kim, Kundu, Viollet, & Guan, 2011). We found a large 50-fold decrease in abundance of the phosphopeptide AS*PHDVLETIFVR (Ser361) between day 3 and day 4 of the culture. This protein was not found to be differentially expressed

in the non-enriched proteomic analysis. For the onset of autophagy, ATG13 is phosphorylated by ULK1 (Alers, Loffler, Wesselborg, & Stork, 2012). Here, we identified 6 phosphopeptides from ULK1 protein from which phosphopeptides containing the identified phosphosites Ser556 on LHS*APNLSDFHVVRPK and Thr670 on NRT*LPDLSEAGPFQGQQLGSGLRPAEDTR show differential expression between day 2 and day 3 of culture with a net decrease in phosphopeptide abundance of 34 and 11-fold respectively. The phosphopeptide SGSTS*PLGFAR containing Ser469 showed a 31-fold decrease from day 3 to day 4 of culture. Three other phosphopeptides containing the phosphosites Ser714 (AAFGTQASDS*GSTDSLQEKPMEIAPSAGFGGTLHPGAR), Thr452 (IEQNLQSPT*QHQTAR), and Ser477 (AS*PSPPSHTDGAMLAR) were found to show decreased expression from day 2 to day 5 of culture.

Several other phosphoproteins with potential implication in recombinant protein productivity and product quality were identified to be differentially expressed between the growth phases. For example, we found four differentially expressed phosphopeptides from the intermediate filament protein, vimentin, during culture progression. Previous proteomic studies have shown that increased expression of vimentin is associated with high specific productivity (Meleady, 2007) (Kumar, Gammell, Meleady, Henry, & Clynes, 2008)(Baik et al., 2006). We identified four phosphopeptides with phosphosites Ser55 (SLYSS*SPGGAYVTR), Ser56 (SLYSSS*PGGAYVTR), Ser459 (DGQVINETS*QHHDDLE), Ser412 (LLEGEESRIS*LPLPNFSSLNLR) and Ser325 (QVQS*LTCEVDALK) of vimentin to show decreased expression over the culture duration. We also found differential expression of three phosphopeptides from signal recognition particle subunit (SRP72), an essential protein for targeting secreted protein to the rough endoplasmic reticulum

(M. M. M. Becker, Lapouge, Segnitz, Wild, & Sinning, 2017). SUMOylation, the covalent attachment of SUMO (small ubiquitin-like modifier) of a protein product, has shown to increase product quality by improving folding, solubility, and stability of the recombinant protein (Peroutka, Elshourbagy, Piech, & Butt, 2008). We identified differential expression of a number of phosphoproteins associated with the SUMOylation pathway including SENP1 and NUP153. In high producer cells, the secretory pathway remains a bottleneck (Le Fourn et al., 2014). A recent study showed a dramatic increase in recombinant protein productivity by knockdown of two secretory pathway proteins, i.e. Ceramide Synthase 2 (CERS2) and Tbc1D20 (Le Fourn et al., 2014). We identified a 2.7-fold increase in expression of a phosphopeptide expression from CERS2 from day 2 to day 3 of culture. We also identified differentially expressed phosphopeptides from the heat shock proteins, Hspa5 and Hap90b1; the expression of these two proteins has been previously linked with increased protein productivity (Dorner, Wasley, & Kaufman, 1989).

As the cellular mechanisms controlling CHO cell growth and productivity are intertwined, cyclin-dependent kinases (CDKs) have been shown to be prime targets of interest to control CHO cell growth and improve productivity (Kumar, Gammell, & Clynes, 2007). In our study we identified differential expression of eight CDKs including Cdk6, Cdk5rap2, Cdk2, Cdk17, Cdk16, Cdk14 Cdk12, Cdk1 and Cdk6. Interestingly a previous study has demonstrated that manipulation of signalling pathways using small molecule inhibitors that can directly target CDK 4/6 to arrest cell growth after a few days of culture has been found to improve productivity and quality of a Mab (Du et al., 2015). This suggests the potential to manipulate expression of such target phosphoproteins identified to be differentially expressed during the

growth phases in our study, either through genetic engineering or chemical inhibitor approaches, to control cell growth and potentially improve productivity.

2.7 Conclusion

The data in this study clearly demonstrates that studying the proteome and phosphoproteome adds additional dimensions to the understanding of recombinant CHO cells in bioprocess conditions. These phosphoproteins and the kinases and phosphatases that are involved in their regulation have the potential to be engineering targets in CHO cells. Kinases and phosphatases also have the potential to be chemically manipulated, for example, to control cell growth and increase Mab productivity(Du et al., 2015). From a perspective of future CHO cell signaling research, the identified kinases are of particular interest as to fully understand the signaling dynamics of CHO cells grown in bioprocess-relevant conditions. It is equally important to understand the kinase-substrate relationship to pinpoint the responsible kinases for observed differences in phosphosite abundance and provide the opportunity to manipulate CHO cell behaviour by regulation of endogenous kinases.

Future work will involve quantification of phosphorylation changes using mass spectrometry based approaches such as Parallel Reaction Monitoring (Abelin et al., 2016) due to the lack of availability of CHO-specific anti-phospho antibodies. Genome editing techniques such as CRISPR/Cas9 are also ideal tools to study the phosphoproteome in further detail and will allow the assignment of functionality to specific phosphosites of interest through targeted site-specific mutagenesis approaches (Lee et al., 2015).

Supplementary data
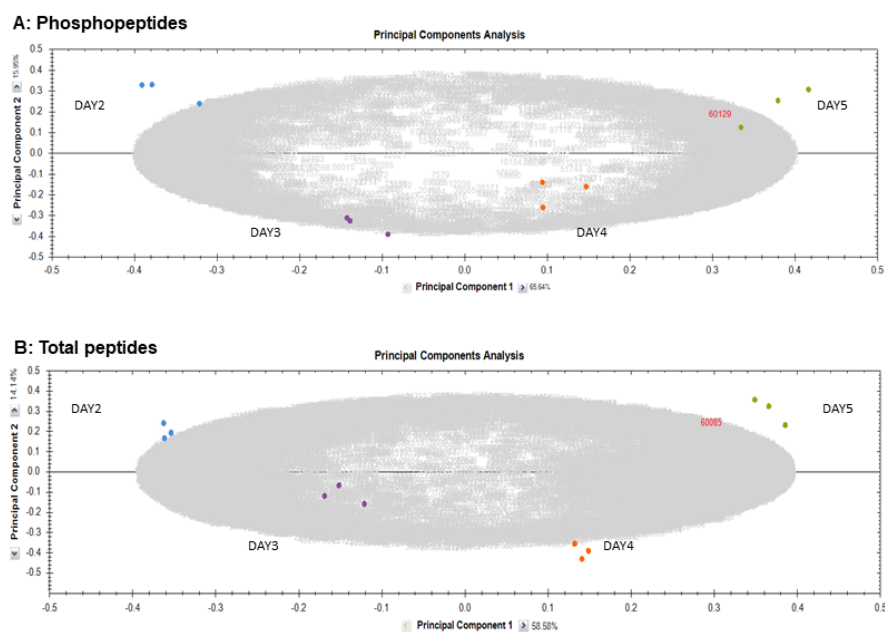
Supplementary Figure 1 PCA plot output from Progenesis QI for Proteomics showing clustering of differentially expressed (A) phosphopeptides and (B) peptides from the different time points

Supplementary tables can be downloaded from

https://drive.google.com/open?id=1XskMA0UiS2oHTOZkaGyAob65nbUihR_G
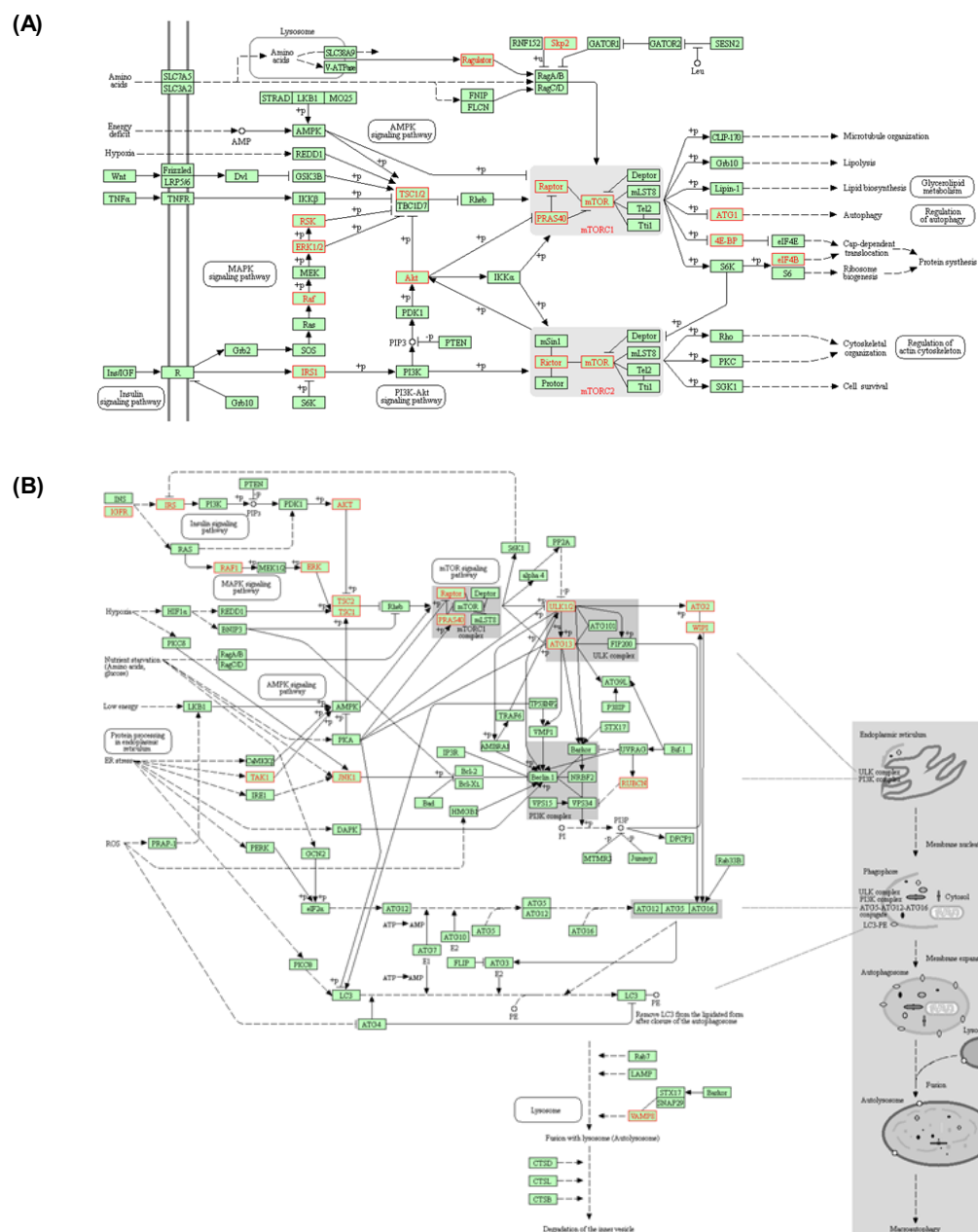
Supplementary Figure 2 Gene Ontology: KEGG Pathway; (A) The mTOR pathway and (B) the autophagy pathway. The pathway component proteins for which differentially expressed phosphopeptides were identified are highlighted in red.

3. Chapter 3 Increased mAb production in amplified CHO cell lines is associated with increased interaction of CREB1 with transgene promoter

The work presented in this chapter was performed in collaboration with visiting Fulbright scholar Prof Susan T.Sharfstein from SUNY polytechnic, and the resultant manuscript was prepared as co-first authors between Dr Hussain Dahodwala and Myself. I contributed towards analysis of RNA-seq data, proteomic sample preparation, differential proteomics, and analysis of proteomics and phosphoproteomics results, data visualization and discussion, manuscript draft writing and carried out subsequent revisions.

Highlights

- Cell line selection and MTX adaptation increase transgene expression in industrial CHO cells

- RNA-Seq and bioinformatic studies implicate transcription factors in increased production.

- Changes in nuclear proteome and phosphoproteome are evident in MTX-amplified cell clone.

- Chromatin immunoprecipitation demonstrates an increased association of CREB1 with the transgene.

Abstract

Most therapeutic monoclonal antibodies in biopharmaceutical processes are produced in Chinese hamster ovary (CHO) cells. Technological advances have rendered the selection procedure for higher producers a robust protocol. However, information on molecular mechanisms that impart the property of hyper-productivity in the final selected clones is currently lacking. In this study, an IgG-producing industrial cell line and its methotrexate (MTX)-amplified progeny cell line were analyzed using transcriptomic, proteomic, phosphoproteomic, and chromatin immunoprecipitation (ChIP) techniques. Computational prediction of transcription factor binding to the transgene cytomegalovirus (CMV) promoter by the Transcription Element Search System and upstream regulator analysis of the differential transcriptomic data suggested increased *in vivo* CMV promoter-cAMP response element binding protein (CREB1) interaction in the higher producing cell line. Differential nuclear proteomic analysis detected 1.3-fold less CREB1 in the nucleus of the high productivity cell line compared with the parental cell line. However, the differential abundance of multiple CREB1 phosphopeptides suggested an increase in CREB1 activity in the higher producing cell line, which was confirmed by increased association of the CMV promotor with CREB1 in the high producer cell line. Thus, we show here that the nuclear proteome and phosphoproteome have an important role in regulating final productivity of recombinant proteins from CHO cells, and that CREB1 may play a role in transcriptional enhancement. Moreover, CREB1 phosphosites may be potential targets for cell engineering for increased productivity.

Keywords: Chromatin immunoprecipitation (ChIP), CHO cell line selection, nuclear proteomics, transcriptional regulation, Cyclic AMP-responsive element-binding protein (CREB)

3.1 Introduction

The development of hybridoma technology by Kohler and Milstein (Köhler & Milstein, 1975), set the stage for the development of monoclonal antibodies (mAbs) as tools in research, diagnostic agents, and revolutionary therapeutic agents, treating a wide range of indications. Chinese hamster ovary (CHO) cells have emerged as the dominant host for production of protein biopharmaceuticals, particularly monoclonal antibodies. As increasing numbers of therapeutic protein candidates enter various stages of development, biopharmaceutical companies are seeking innovative solutions to deliver this pipeline (Tejwani et al., 2018). Therefore, in this competitive market, it is essential to reduce time to market while maintaining desired quality attributes. Moreover, due to the large doses of antibody therapeutics required over an extended period for some indications, manufacturing capacity becomes an issue. To meet the high demand for biopharmaceuticals, many companies have built large-scale manufacturing plants containing multiple 10,000 L or larger cell-culture bioreactors. In this paradigm, high-producing cell lines significantly impact the development timelines and reduce costs by reducing needed bioreactor capacity and process cycles. Traditional cloning methods for cell line production and selection have many shortcomings and are labor intensive and time consuming. Even with the development of platform technologies and processes, each biopharmaceutical molecule still requires labor-intensive clone selection. Currently, there is a lack of understanding of the cellular organization and mechanism of high productivity, hindering the rapid development and selection of higher producing clones.

Cell line development is currently performed in the following steps:

1. A host cell line is transfected with a transgene-harboring plasmid via an optimized protocol. Flow cytometry-based staining techniques are frequently used to confirm transfected clones.
2. Pools are amplified and selected using a chemical reagent.
3. Single cells are isolated, scaled up, and adapted.
4. A final clone is selected based on titer and stability.

A top clone is not merely isolated from a pool of differential producers, rather the cell line adaptation to screening pressures results in genomic and phenotypic changes that gives rise to the final top clone (Noh, Shin, & Lee, 2018). Methotrexate (MTX) amplification is routinely employed in dihydrofolate reductase-negative (DHFR$^-$) systems to select for higher producers, with similar amplification performed using methionine sulfoximine (MSX) in glutamine synthetase deficient (GS$^-$) systems. We previously characterized various CHO cell clones producing the same recombinant humanized monoclonal antibody and observed that MTX amplification leads to increased productivity by not only causing an increase in transgene copy number but also by transcriptional enhancement in higher producer CHO cell lines (Jiang et al., 2006). Thus, in these clones, the process of transcription is the rate-limiting step in recombinant antibody production. Further work demonstrated that treatment with sodium butyrate can improve gene expression in these clones (Jiang & Sharfstein, 2008). Sodium butyrate is a known histone deacetylase inhibitor (Yin et al., 2018) and improved productivity may occur by increasing the accessibility of transgene to the transcriptional machinery. Therefore, productivity can potentially be improved by altering the DNA-protein interactions in the cells.

The molecular basis for maximal expression from a defined section of DNA is dependent on the state of the chromatin. Changes in gene expression are governed by factors outside the realm of sequence information (Dahodwala & Sharfstein, 2014).

These epigenetic changes are cell-type specific (Feichtinger et al., 2016) (Akopov et al., 2006). Based upon epigenetic mechanisms, many strategies have been devised both to generate stably transfected clones as well as to increase specific productivity (Dahodwala & Sharfstein, 2017b). While there has been considerable success in exploiting these observations to improve specific productivity, there is no clear understanding of the role of the transcriptional proteins involved. Recent computational and experimental studies exploring the interactions of transcription factors with the cytomegalovirus (CMV) promoter in the context of transient transfection and production of the reporter proteins secreted alkaline phosphatase and green fluorescent protein identified several transcription factor regulatory elements in the CMV promoter that affected transcription, particularly the cAMP response element (CRE) and nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) as positive regulatory elements and the binding site for the zinc finger regulatory protein YY1 as a negative regulatory site (Brown, Sweeney, Mainwaring, & James, 2014, 2015).

While these previous studies provide insight into potential interactions of transcription factors with the CMV promoter in CHO cells, they do not address the potential changes in transcription factor regulation during cell line selection for stable clones or chromatin modification during stable incorporation of transgenes into the host chromosomes. For example, we previously observed variability in metabolic behavior between different clones, all from the same host cell line, producing the same recombinant product, grown in the same medium under the same culture conditions, presumably as a result of modifications that occurred during the selection process (Dahodwala, Nowey, Mitina, & Sharfstein, 2012). Whether this occurs as a result of stress from increased productivity, increasing levels of MTX or other factors remains

to be elucidated. In the mutable genome of the CHO cells, the changes in chromatin and nuclear proteome resulting from such adaptations will have a profound effect on the mechanism of productivity of the derived clones. Changes in protein expression and post-translational changes such as phosphorylation lead to nuclear translocation of transcription factors and subsequent changes in DNA binding (Kaushik et al., 2018). Many cofactors themselves may exhibit histone acetylation activities, thereby modifying the chromatin accessibility and subsequent gene regulation (Zupkovitz et al., 2006). These observations indicate that transgene expression may be affected by inherent differences in levels and modifications of transcription-factor binding proteins and their subsequent interaction with the promoters in different cell lines. In this study, comparative phosphoproteomic data were gathered from a mAb-producing clone (A0) and its MTX-amplified progeny (A1), using quantitative, label-free LC-MS proteomic techniques to demonstrate the activation and increased phosphorylation of CREB1 in the amplified cell line. Further, chromatin-transcription factor interactions were investigated by comparing the parental clone and its MTX-amplified progeny via chromatin immunoprecipitation (ChIP). An increased DNA-protein interaction in the higher producing cell lines was observed. CREB1 transcription factor showed ~6-fold increased association with the cytomegalovirus (CMV) promoter in higher-producing cell lines. Together these results indicate an increasing association of transcriptional proteins with the DNA in the higher producing clones, reinforcing the notion that epigenetics and nuclear proteome interplay is an important, but poorly understood driver of transgene expression in mammalian cells.

3.2 Methods and Materials

3.2.1 Cell lines

Chinese hamster ovary cell lines that produce a recombinant monoclonal humanized IgG with different specific productivities were a generous gift from an industrial collaborator. These cell lines were developed by co-transfecting two plasmids, one containing IgG heavy chain (HC) and dihydrofolate reductase (DHFR) genes and the other containing IgG light chain (LC) and neomycin phosphotransferase (Neo) genes. Transfected cell lines were initially selected in medium containing 400 μg/mL neomycin (G418). After selection, the neomycin was removed, and all subsequent cultures were performed in the absence of neomycin. Subsequently, gene amplification was performed by stepwise selection with increasing MTX concentrations. For these studies, a low producing parental cell line A0 and its amplified high-producing progeny cell line A1 were chosen for investigation. These cell lines have been previously described (Jiang & Sharfstein, 2009) After culture medium adaptation, cells were cultured in a nonproprietary, serum-free medium ((Dahodwala et al., 2012) with hydrolysate, 5 mg/L recombinant human insulin and MTX (Supplementary Table C3-3). Growth curves and relative specific productivities are shown in Supplementary Figure 4.

3.2.2 Cell culture conditions

For every experimental method described, triplicate batch suspension cultures of all cell lines were maintained in 125 mL Erlenmeyer flasks. For each cell line, 0.2×10^6 cells were seeded into 25 mL of medium and cultured on an orbital shaker at 125 rpm, 36° C, and 5% CO_2. Routine subculturing was carried out for 2 passages after thawing before experiments were performed.

3.2.3 Sampling

Samples were taken daily from suspension cultures to determine cell density and viability. Cell densities and viabilities were estimated by hemacytometer counts (Hausser Scientific, PA) or automated counting (BioRad TC10) after diluting 1:1 with 0.4% trypan blue solution.

Cell pellets were collected at mid-exponential stage in culture (3 days after inoculation). Cells were harvested by centrifuging the appropriate volume of culture suspension at 1200 rpm for 5 min.

3.2.4 Antibody Assay

Antibody titers were determined by ELISA using a Human IgG ELISA Antibody Pair Kit (Stemcell Technologies) as per the manufacturer's instructions.

3.2.5 Antibodies

Antibodies to RNA polymerase II were provided with the ChIP IT kit (Active motif, Carlsbad CA). Antibodies to CREB1 (39013), NFκB (40916) and Sp1 (39058) were purchased separately from Active Motif, Carlsbad, CA.

3.2.6 Prediction of transcriptional proteins interacting with CMV promoter region

The Transcription Element Search System (TESS) previously available at http://www.cbil.upenn.edu/cgi-bin/tess/tess is a database that contains the various binding consensus sequences that are recorded by experimental investigation of the transcriptional proteins (Schug, 2008). It can identify binding sites using site or consensus strings and positional weight matrices from the TRANSFAC, JASPAR, IMD, and the CBIL-GibbsMat database. By querying the database with the CMV promoter sequence, we were able to generate a probability score of CMV promoter-region interactions with all transcription factor proteins in the database.

3.2.7 RNA-seq analysis

RNA-Seq data were generated as previously reported (Chiang et al., 2019). Briefly, total RNA was isolated using a Qiagen RNeasy Plus Mini Kit as per the manufacturer's instructions. RNA quality was verified using an Agilent Bioanalyzer prior to library preparation. Library preparation was performed with an Illumina TruSeq Stranded mRNA Library Prep Kit High Throughput (Catalog ID: RS-122-2103), according to manufacturer's protocol. Final RNA libraries were first quantified by Qubit HS and then QC on Fragment Analyzer (from Advanced Analytical). The final pool of libraries was analysed on the Illumina NextSeq platform with high output flow cell configuration (NextSeq® 500/550 High Output Kit v2 (300 cycles) FC-404-2004).

3.2.8 RNA-Seq data processing

The RNA libraries were mapped to the CHO genome (C_griseus_v1.0) (N. E. Lewis et al., 2013a; X. Xu et al., 2011a) using STAR aligner (v. 2.5.4b) (Dobin et al., 2013). Alignments were processed to quantify gene expression counts with HTSeq-count (v. 0.7.2) (Anders, Pyl, & Huber, 2015). Genes with very low expression (less than one count in at least two samples) or of zero variance were excluded from further downstream analysis. DESeq2 with default parameters (Love, Huber, & Anders, 2014) was used to estimate the differential expression between the A1 and A0 samples, with a positive fold change denoting higher expression in A1. The raw sequencing files and count matrix were deposited to SRA and GEO (accession number GSE13351). To comply with intellectual property requirements, the sequencing data were processed to exclude unmapped reads. This results in less than 5% reduction in available reads.

3.2.9 Nuclear proteomics and phosphoproteomics

For proteomic analysis, cells from three biological replicates per condition were harvested at the mid-exponential phase of the culture. The nuclear proteomic fractions were enriched using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific – 78833) as per the manufacturer's guidelines. Protein quantification was carried out using Quick Start Bradford protein Assay (BioRad). To prepare the samples for mass spectrometry analysis, 1 mg of protein lysate from each sample was reduced by adding dithiothreitol to a final concentration of 5 mM and incubated at 56 ºC for 25 min. Samples were then alkylated by adding iodoacetamide to a final concentration of 14 mM and incubated for 30 min at room temperature in the dark. Alkylated samples were then vortexed and diluted at a ratio of 1:5 in 25 mM Tris-HCl. Protein samples were subsequently digested using trypsin (MS grade, Thermo Fisher Scientific) at 1:50 enzyme: substrate ratio. After a 4 hr initial incubation at 37 °C, a further addition of trypsin at 1:100 enzyme: substrate ratio was performed followed by overnight incubation. After overnight digestion, trifluoroacetic acid (TFA) was added to each sample to a final concentration of 0.4% to inactivate trypsin. Peptides from the digested protein were concentrated and desalted using Sep-Pak C-18 columns with negative pressure (Villén & Gygi, 2008). Ten percent of the eluate was aliquoted for total proteome analysis. The remaining 90 % was used for phosphopeptide enrichment using Fe-NTA (IMAC) spin columns (Pierce, Thermo Fisher Scientific) as per manufacturer's instructions. Non-enriched peptide and phosphopeptide sample concentrations were determined using a Nanodrop One (Laptech International, UK).

3.2.10 LC-MS/MS Analysis

Both enriched phosphopeptide samples and peptides previously collected for total proteomic analysis were dried in a SpeedVac vacuum concentrator and resuspended

in 0.1% formic acid (FA) containing 2% acetonitrile (ACN). Peptide volume equivalent to 1 µg total protein was injected by autosampler for LC-MS/MS analysis using an UltiMate 3000 nanoRSLC system (Thermo Scientific) coupled in-line with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific). Prior to the nanoLC separation, samples were first loaded onto the trapping column (PepMap100, C18, 300 µm × 5 mm) for 3 min at a flow rate of 25 µL/min with 2% (v/ v) ACN, 0.1% (v/v) TFA. The trapped peptides were back-flushed onto the analytical column (Easy-Spray C18 75 µm × 250 mm, 2 µm bead diameter column) using a gradient of 98% A (0.1% (v/v) FA): 2% B (80% (v/v) ACN, 0.08% (v/v) FA) to 35% B over 120 min at a flow rate of 300 nL/min.

Data-dependent product ion mode was applied for both non-enriched and phosphopeptide-enriched MS analysis. For peptide precursor fragmentation and detection, the full MS survey scan (m/z 380-1500) was performed at a resolution of 120,000 with the automatic gain control (AGC) target set to 5×10^5. Peptides with charge states between 2 and 7 were selected for MS/MS with the instrument running in top speed mode with a cycle time of 3s. Dynamic exclusion was enabled with the repeat count set to 1, exclusion duration set to 60 s and a mass tolerance of +/- 10 ppm. For non-enriched peptide samples, MS^2 was performed following quadrupole isolation with HCD fragmentation using normalized collision energy of 28% in the ion trap (IT). MS^2 spectra were acquired with a fixed first m/z of 100 and an intensity threshold of 5,000. AGC was set to accumulate 1×10^4 ions and the maximum injection time was 35 ms.

For phosphopeptide-enriched peptide samples, multistage activation (MSA) was performed following quadrupole isolation for CID fragmentation with the normalized collision energy set to 32%, CID activation time of 10 ms and activation Q of 0.25 in

the IT. An intensity threshold of 10,000 was used. The neutral loss mass for MSA was 97.9673, AGC was set to accumulate 2×10^4 ions and the maximum injection time was 90 ms.

3.2.11 Quantitative Label-free LC-MS/MS Analysis

Relative quantitative label-free LC-MS analysis of the total proteome and phosphoproteome fractions from the two cell lines was carried out using Progenesis QI for Proteomics (Nonlinear Dynamics, Waters) in conjunction with Proteome Discoverer 2.2 (Thermo Scientific) for protein identification utilizing Sequest HT (Eng, McCormack, et al., 1994) search algorithm as previously described (Henry, Power, et al., 2017a). Raw files generated from the MS/MS analysis were imported into Progenesis QI, and automatic reference alignment was carried out to account for retention time variability between LC runs. Upon alignment of all runs, identified features were filtered based on ANOVA p-value < 0.05 between experimental groups. For proteomic and phosphoproteomic analysis, a Mascot generic file (mgf) was generated from all exported MS/MS spectra and analyzed using Proteome Discoverer v.2.2 (Thermo Fisher Scientific) in conjunction with SEQUEST. Peak lists were searched against a proteogenomic draft annotation for the newly assembled Chinese hamster genome which is experimentally annotated using RNA-Seq, proteomics, and Ribo-Seq (S. Li et al., 2019). Database search parameters were set to allow MS1 tolerance of 10 ppm; MS^2 mass tolerance of 0.6 Da for ion trap detection; enzyme specificity was set as trypsin with two missed cleavages allowed; carbamidomethylation of cysteine was set as a fixed modification; and phosphorylation of serine, threonine, and tyrosine and oxidation of methionine (for phosphopeptide analysis) were set as variable modifications. For phosphosite identification, the PhosphoRS algorithm (Potel et al., 2019) was run through

ProteomeDiscover 2.2 using diagnostic fragment ions and analyzer-specific fragment ion tolerances, as described above. Data were filtered to a 1% false discovery rate (FDR) on PSMs using automatic decoy searching in SEQUEST and by applying a phosphosite probability score of 75% or greater for S, T or Y amino acids in PhosphoRS (Potel et al., 2019). A statistical criterion of ANOVA p-value ≤ 0.05 and fold change cut-off ≥ 1.5 at the protein level was applied between experimental groups. Proteins with two or more unique peptides and phosphoproteins with any unique phosphopeptides that passed these criteria were then deemed to be differentially expressed between the relevant experimental groups being analyzed.

3.2.12 Purification of the DNA-protein complex

Chromatin immunoprecipitation was performed using the ChIP-IT kit (53008 Active Motif, Carlsbad CA) according to the manufacturer's instructions. Briefly, 4×10^7 cells from each cell line were harvested at day 3 and incubated with 30 mL fresh medium containing 1.5 mL 36 % formaldehyde (47630 Sigma-Aldrich, St Louis, MO) for 10 min to crosslink the DNA-associated proteins to the chromatin. The reaction was stopped by washing the cells with phosphate buffered saline (PBS) and incubating with Glycine Stop-fix solution for 10 min. A final PBS wash step was used to clean the cell pellet. A sonicator (450D Branson, Danbury CT) fitted with a microtip was employed to disrupt the cells and shear the DNA to 500-1500-base pair fragments. The sonicator settings were set in accordance with the tip manufacturer's instructions and kept at 40% amplitude. The shearing was verified by separating the sheared DNA on a 1.8 % agarose gel. In subsequent steps, the Protein-DNA complex was immunoprecipitated using antibodies to CREB1, NFκB, Sp1 or RNA polymerase II. After treatment with Proteinase K and RNase to remove cellular proteins and RNA, DNA fragments were purified by using silica spin columns provided with the kit. The

final elution volume in each fraction was 100 µL. This volume was concentrated to 20 µL by using a SpeedVac DNA concentrator (BC-SDNA11 Savant, GMI Inc. Ramsey, Minnesota) to obtain an adequate concentration of DNA template for RT-qPCR. ChIP was performed on three separate dates with duplicate PCR analysis for each sample.

3.2.13 RT-qPCR

Real time quantitative PCR was performed using the Roche *LightCycler®* 480 Real-Time PCR System and the LightCycler 480 Mastermix (04707494001 Roche, Indianapolis, IN). For quantification of CMV, the probe/primers combinations were as follows: forward primer: gcagagctcgtttagtgaacc; reverse primer: gaggtcaaaacagcgtggat; Universal ProbeLibrary probe: #80 (cat.no. 04689038001, Roche, Indianapolis, IN). For quantification of glyceraldehyde 3-phosphate dehydrogenase (GAPDH), the probe/primers combinations were as follows: forward primer: cgtattggacgcctggttac; reverse primer: ggcaacaacttccactttgc; Universal ProbeLibrary probe: #8 (cat.no. 04685067001, Roche, Indianapolis, IN). Reaction conditions were set up according to the manufacturer's instructions. Crossing points (Ct) were generated from the LightCycler Software. Relative quantification of the CMV promoter and GAPDH bound to the transcription factors was performed using the 2^delta delta Ct method (Rao, Huang, Zhou, & Lin, 2013). All samples were normalized to the respective input DNA for the ChIP reaction (e.g. A0 cell line, CMV copies in input DNA) and then to sample 3 of the A0 CREB1 precipitate for CMV or GAPDH, respectively.

3.3   Results

3.3.1   Transcription factor-binding analysis of CMV promoter

In this study, protein-DNA complex interactions were examined to understand the transcriptional enhancement in the MTX-amplified cell lines and further elucidate

transcriptional regulation in high productivity clones. Transcription factors associate upstream of the gene of interest via specific binding motifs that interact with consensus sequences along the promoter region to initiate transcription.

The TESS web tool was used for predicting transcription factor binding sites in the CMV promoter sequence. The CMV promoter is a strong viral promoter system used in transgene expression. It has a high level of constitutive gene expression and is efficient in a broad range of cell types. Putative transcription factors with a high probability of binding included enhancer factor C (EF-C), methylated DNA-binding protein (MDBP), activator protein 1 (AP-1), nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB), activating transcription factors (ATF), cAMP response element binding protein (CREB1), and activating protein 2 (AP-2) as shown in Table 3.1.

Table 3.1: Top scoring transcription factors binding sites 10 out of 548 sites are shown. From TESS: Transcription Element Search System (http://www.cbil.upenn.edu/cgi-bin/tess/t) along with the z-score values from Ingenuity URA analysis.

#	Transcription Factor	Beg	Sns	Len	Sequence	*La	**URA Z-Score
1	T00230 EF-C T00963 MDBP	292	R	25	TATGTTCCCATAGT AACGCCAATAG	50.0	
2	T00029 AP-1 T00590 NF-κB	562	R	23	TTTGACTCACGGGG ATTTCCAAG	46.0	1.74
3	T00051 ATF T00167 CREB 1	592	R	18	CCCATTGACGTCAA TGGG	36.0	2.135
4	T00952 AP-2	99	N	8	CCCMNSSS	10.0	

*La: Log likelihood score of association of transcription factor to given consensus sequence

**URA Z score: Likelihood of activation (positive Z-score) or inactivation (negative Z-score) of transcription factor based on differential expression of genes regulated by the transcription factor

3.3.2 RNA-Seq - upstream regulator analysis

Differential gene expression analysis between the A0 and A1 cell lines was performed from the count data obtained from RNA-Seq analysis using DEseq2 package in R. 6272 genes were identified as differentially expressed between the A1 and A0 cell lines (p_{adj} <0.05) from which 2938 genes exhibited higher expression in the higher producer progeny A1 cell line, whereas 3334 genes showed lower expression in A1 as compared to the parental A0 cell line.

To understand key genes that regulate productivity associated pathways, Ingenuity Upstream Regulator Analysis (URA) was performed on the list of differentially expressed genes. URA analysis provides a numerical account of predicted regulator activity as Z-score value. A positive Z-score > 1.5 indicates potential activation of a regulatory factor and Z-Score <-1.5 indicates potential inhibition of the regulatory factor (Krämer, Green, Pollard Jr, & Tugendreich, 2013). We found 102 regulatory factors with Z-score > 1.5 and 56 regulatory factors with Z-score < -1.5. Among all predicted regulatory proteins, 28 transcription factors were predicted to be activated, and 11 transcription factors were predicted to be inhibited in the high producer A1 cell line. Examining the transcription factors with a high probability of interaction with the CMV promoter, both CREB1 and NFκB were identified by the URA as having a high probability of being activated in the A1 cell line compared with the A0 parental cell line. Although no expression change for CREB1 was observed at the mRNA level, URA prediction suggested strong activation of CREB1 (z-score = 2.13). In addition, the NFκB complex was identified as likely to be activated, with a z-score = 1.7. URA analysis also predicted potential activation of 11 kinases and inhibition of 6 kinases in the A1 cells, an important class of regulatory proteins that influence transcription

factor activation. A complete list of all transcription factors with predicted differential activation is provided in Figure 3.1.
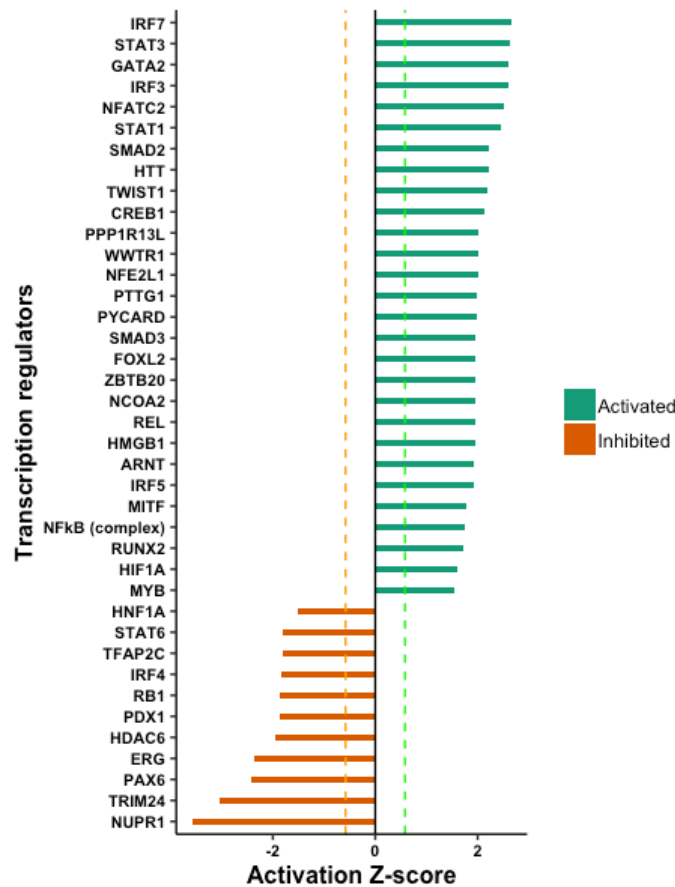
Figure 3.1: Ingenuity pathway analysis-Predicted activity status of transcription regulators from upstream pathway analysis

3.3.3 Proteomic and phosphoproteomic changes associated with high productivity

Previous proteomic studies have shown that changes in protein productivity in CHO cells bring about systemic changes in the proteome (Hausmann, Chudobová, Spiegel, & Schillberg, 2018) and post-translation modification of cellular proteins (Henry, Power, et al., 2017b). Due to the compartmentalization of transcriptional machinery in the nucleus, studying the nuclear proteome can provide insight into changes in expression and modification of proteins that influence transgene expression. Furthermore, nuclear proteomic studies can provide up 60% greater coverage of key

regulatory proteins such as transcription factors, transcription co-regulators and RNA processing proteins (J. Wang et al., 2017) than whole cell proteomic analysis.

Quantitative nuclear proteomic and nuclear phosphoproteomic analysis of high producer A1 cell line and low producer A0 cell line revealed 873 differentially expressed proteins (DEP) and 640 differential phosphoproteins (DEpP). From the DEP, 426 proteins were elevated in the high producer A1 cell line and 447 proteins were more highly expressed in the low producer A0 cells. Similarly, from the nuclear phosphoproteomic analysis, 528 phosphoproteins showed elevated phosphorylation and/or increased expression and 112 phosphoproteins showed reduced phosphorylation and/or decreased expression in the A1 cell line compared to the A0 parental cell line. Only proteins and phosphoproteins with fold change ≥ 1.5 were deemed differential. 178 proteins were common between the list of differential proteins and differential phosphoproteins. A complete list of all nuclear proteins, phosphoproteins, and phosphopeptides is presented in Supplementary Table C3-1.

Although no differential mRNA expression for CREB1 was identified in the transcriptomic data, 1.3-fold lower expression of total CREB1 protein was observed in the high producer A1 cell line compared with the A0 cell line (**Error! Reference source not found.**A, Supplementary Table C3-1). In the light of TESS prediction and URA analysis, the observed decrease in CREB1 total protein in the nucleus of the A1 cell line could be due to the increased activity of CREB1 in A1 cells, leading to greater CREB1-DNA association, decreasing the availability of free CREB1 in the nuclear proteome for proteomic detection. Moreover, phosphoproteomic comparison between A1 and A0 cell line revealed differential abundance of three phosphopeptides of CREB1 corresponding to four phosphosites, i.e. ILNDLsSDAPGVPR (Ser-148), RLFSGTQISTIAESEDsQEsVDSVTDSQK (Ser-111, Ser-114) and

TAPTSTIAPGVVMASsPALPTQPAEEAAR (Ser-271). The expression pattern of these phosphopeptides is shown in Figure 3.2 B-D (See also Supplementary Table C3-1).
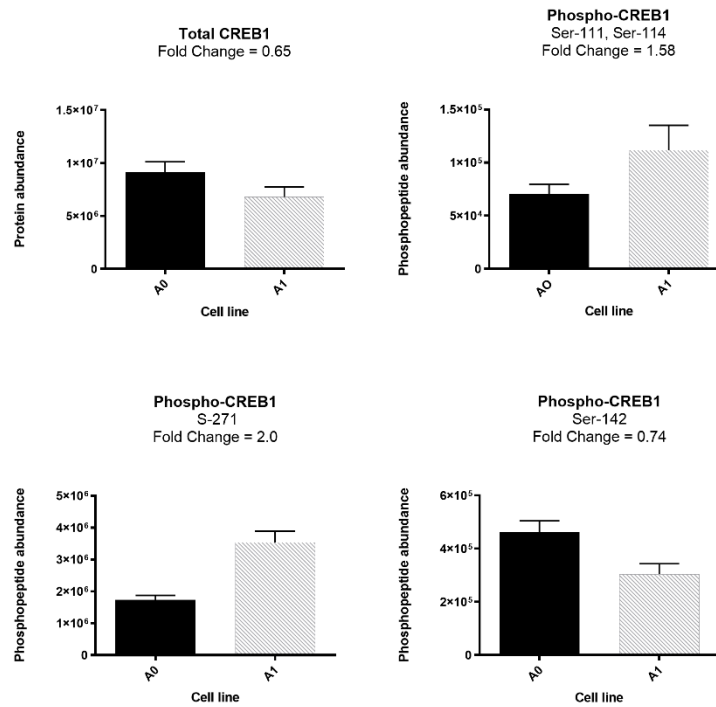
Figure 3.2: CREB1 proteomic and phosphoproteomic expression changes in the nucleus of A0 and A1 cell lines. A-Total CREB1, B-D-phosphopeptides Ser-111 and Ser-114, Ser-272, and Ser-142, respectively. Fold change is the relative abundance in A1 relative to A0.

3.3.4 CREB1: Transcription regulatory relationship

The expression pattern of transcription-factor target proteins can provide insight into the dynamics of regulation by the transcription factor. Hence, CREB1-interacting proteins from TRRUST (Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining) database were searched in the lists of differential genes, proteins, and phosphoproteins between the A0 and A1 cell lines. The expression pattern of the six CREB1 target proteins identified is shown in Table 3.2. Two proteins that are repressed by CREB1 activation, HMOX1 and JUN, were downregulated at the mRNA and protein level, and protein JUN also showed downregulation at the phosphoprotein level in the A1 cell line. Both proteins showed a high likelihood of

inhibition in the URA analysis. Two proteins that are upregulated by CREB1, NOLC1 and NDC80, were upregulated at the total protein level in the A1 cell line. Protein NOLC1 was also upregulated at the phosphoprotein level whereas the mRNA level for NOLC1 was downregulated in the A1 cell line.

Gene	Activity	RNA-seq	Total Protein	Phosphoprotein	URA-ZSCORE
HMOX1	Repression	-0.70	-1.47	N/A	-2.07
JUN	Repression	-0.42	-1.27	-0.18	-0.34
XPC	Unknown	N/A	-0.76	N/A	N/A
NOLC1	Activation	-0.33	0.89	1.40	N/A
NDC80	Activation	N/A	2.85	N/A	N/A
ETV3	Activation	0.21	N/A	0.61	N/A

Table 3.2: CREB1 regulatory relationship proteins from TRRUST (https://www.grnpedia.org/trrust/) transcriptional regulatory database. Fold change values are shown as Log$_2$FC(A1/A0). Orange cells represent downregulation and green cells represent upregulation in the A1 cell line.

3.3.5 Chromatin immunoprecipitation

Previous work in our lab has identified transcription as the rate-limiting step in the production of monoclonal antibodies in both the parental cells and amplified progeny in these cell lines (Jiang et al., 2006). Based upon the consensus binding sequences of the transcriptional proteins and their occurrences along the CMV promoter, bioinformatic analysis of the CMV sequence indicated that CREB1 exhibits a high probability of influencing transcription from the CMV promoter (Table 1). The URA and nuclear proteomic and phosphoproteomic analysis further substantiated the potential role of CREB1. The URA also indicated a high likelihood of NFκB activation; however, we were unable to find any additional proteomic or

phosphoproteomic evidence to support this activation. To evaluate whether high and low productivity clones exhibited differential interactions between the CMV promoter region and these transcription factors, chromatin immunoprecipitation was carried out, employing antibodies to these transcriptional proteins as well as to Sp1, a negative control.

Immunoprecipitated chromatin was purified and then subjected to RT-qPCR to quantify the number of copies of GAPH and CMV promoter regions bound to each transcription factor. After immunoprecipitation, PCR-derived Ct values were normalized to the respective input DNA values to serve as a loading control. We observed a ~6-fold difference in the CREB1 association with the CMV promoter chromatin in the amplified progeny A1 relative to the A0 parental cell line, indicating that the CREB1 association to the CMV promoter was increased even after accounting for MTX amplification and subsequent increase in CMV copies (Figure 3.3). The CMV binding to NFκB and RNA pol II was much lower than the binding to CREB1. As expected, the CMV binding to Sp1 was quite low as the Sp1 transcription factor has a predicted low likelihood of association with the CMV promoter.

When the immunoprecipitated DNA was assayed with GAPDH primer probes (Figure 3.4), we observed a ~4-fold increase in the GAPDH-CREB1 association in the amplified clone A1 relative to the progeny clone A0. These data suggest that the transcription factor CREB1 shows an increased association with the genomic DNA within the A1 amplified progeny relative to the parental A0 clone, which would be consistent with activation of CREB1 in the amplified cell lines. Despite the predicted increase in NFκB activity from the URA, no increased association of NFκB or RNA pol II with either CMV or GAPDH was observed.
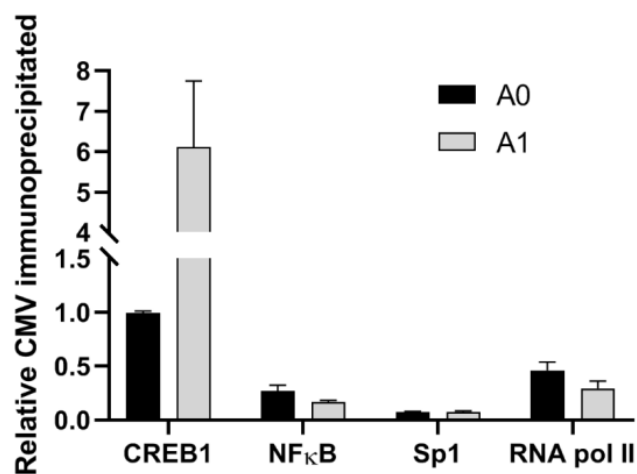
Figure 3.3 Normalized number of copies of CMV promoter region binding to transcription factors in A1 cell line vs. A0 cell line
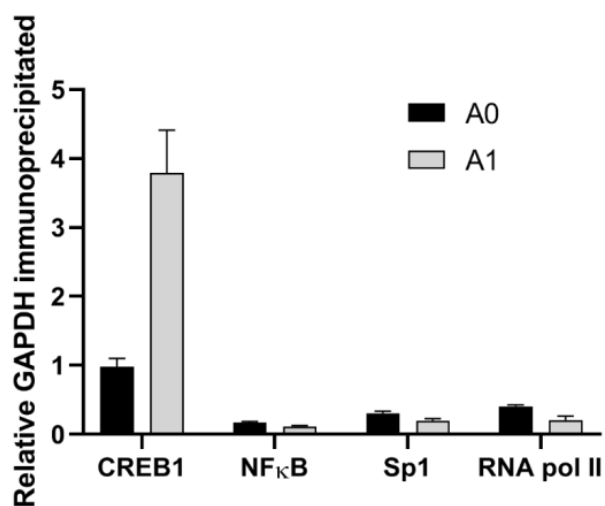

Figure 3.4 Normalized number of copies of GAPDH promoter region binding to transcription factors in A1 vs. A0 cell line

3.4 Discussion

3.4.1 Role of CREB1 in transcriptional enhancement

In previous studies (Jiang et al., 2006), we observed that the A1 cell line has approximately twice the number of copies of the transgenes but expresses 5-fold more transcripts than the A0 parental clone. Results presented here suggest increased association of CREB1 with the nucleus in the high producer cell line A1 compared with the parental A0 clone. CREB1 is a well-studied transcriptional activator that promotes gene expression from a variety of promoters. CREB1 becomes activated as part of the cAMP cascade. A stimulus from the environment causes cAMP to activate a protein kinase that in turn translocate to the nucleus and phosphorylates the CREB1 protein. Activated CREB1 then binds to its DNA consensus sequence in many promoters in the cell and activates the transcriptional machinery. The consensus sequence (5'-GTGACGT[AC][AG]-3') is present in many viral and cellular promoters, including the constitutive CMV promoter that is used to drive gene expression in the clones in this study and is widely used for recombinant protein production in the biopharmaceutical industry. (Ortega-Martínez, 2015). Typically, CREB1 binds DNA as a dimer and also acts as a coactivator of transcription by associating with UTF1 (Beausoleil et al., 2004). The CREB-binding protein (CBP) also exhibits histone acetyltransferase (HAT) activity. Hence, CREB1 is suggested to have a role in the activation of transcription by DNA binding as well as modulating the chromatin, specifically by acetylating histones H2B and H4 (Harton, Zika, & Ting, 2001).

While CMV is a strong constitutive promoter, it shows great variability in activity depending on the cell line used for transfection (J. Y. Qin et al., 2010). The CMV promoter also has a large CpG island and it is prone to silencing by methylation. Removing CpG islands from the promoter tends to mitigate gene silencing but does

not prevent changes in expression due to histone modifications (S. C. L. Ho, Koh, Soo, Chao, & Yang, 2016). While changes in transcription can occur from differences in histone modifications, in the present study we found fairly limited changes in the histone deacetylase (HDAC) and related protein mRNA levels (Supplementary Figure 3)and even fewer changes at the proteomic and phosphoproteomic levels (Supplementary Table C3-2) in the clones under study, leading us to focus on other proteomic and phosphoproteomic differences in the amplified clones.

CREB1 is a transcriptional activator, that undergoes complex phosphoregulation by multiple kinases at multiple phosphosites. Ser-133 is the most studied phosphosite on CREB1 (H. Wang et al., 2018). Phosphorylation of Ser-133 by PKA activates CREB1 by increasing its DNA binding activity, promoting the recruitment of the co-activator proteins CBP and p300 (Clark et al., 2015). Although we did not find differential phosphorylation of Ser-133 in our phosphoproteomic data, three other phosphopeptides containing four phosphosites were identified as differentially phosphorylated. Various studies have shown phosphosites other than Ser-133 also play a crucial role in CREB1 activity (Johannessen & Moens, 2007). Hence, it is important to discuss the phosphorylation pattern of these phosphosites to understand the complex mechanism of CREB1 regulation by phosphorylation. For example, we found that phosphopeptide ILNDLsSDAPGVPR (Ser-142) exhibited 1.52-fold higher phosphorylation in the low producer A0 cell line. Phosphorylation of Ser-142 by CamKII inhibits CREB1 transactivation in CV1 mammalian cells by preventing CREB1 dimerization (X. Wu & McMurray, 2001). Furthermore, the differentially abundant phosphopeptide RLFSGTQISTIAESEDsQEsVDSVTDSQK (Ser-111 and, Ser-114) contains two phosphosites associated with the ataxia-telangiectasia-mutated and casein kinase 1 (ATM/CK) cluster, a group of closely spaced and conserved serine

phosphosites at the N-terminal portion of CREB1 (S108, S111, S114, S117 and S121). The phosphorylation of these phosphosites by ATM/CK in response to DNA damage inhibits CREB1 mediated transcription. However, ATM-independent phosphorylation of the ATM/CK cluster positively regulates CREB1-mediated transcription by promoting nuclear translocation of cAMP-regulated transcriptional coactivators (S. H. Kim et al., 2016). Also, in response to genotoxic stress, phosphorylation of Ser-111 of CREB1 by ATM primes phosphorylation of Ser-108, Ser-114, and Ser-117 by CK1 and CK2, ultimately leading to inhibitory phosphorylation of Ser-121 on CREB1 by ATM. However, DNA-damage-independent phosphorylation of the ATK/CK cluster does not promote Ser-121 phosphorylation (Shanware et al., 2010). In our data, we did not find phosphorylation of Ser-121, suggesting potential genotoxic stress-independent phosphorylation of the ATM/CK cluster on CREB1 in the high productivity cell line, A1. Genotoxic stress also promotes inhibitory phosphorylation of Ser-271 on CREB1 by homeodomain-interacting protein kinase 2 (HIPK2) (Trinh, Kim, Chang, Mastrocola, & Tibbetts, 2013). We found two-fold increased abundance of phosphopeptide TAPTSTIAPGVVMAsSPALPTQPAEEAAR (Ser-271) in the high-producer A1 cell line. Paradoxically, studies have shown an increase in transactivation activity of CREB1 in response to Ser-271 phosphorylation via recruitment of CBP and p300 (Sakamoto et al., 2010). Furthermore, homeodomain interacting protein kinase 2 (HIPK2) has been shown to induce phosphorylation of Ser-2361, Ser-2363, Ser-2371, Ser-2376, and Ser-2381 residues on CBP. It has been suggested, however, that HIPK2 enhances the transcriptional activity of CBP by antagonizing the repressive action of cell cycle regulatory domain 1 (CRD1), located between amino acids 977 and 1076 in HPIK2 independent manner (Kovács, Steinmann, Halfon, Magistretti, & Cardinaux, 2015). We observed the differential

abundance of CBP C-terminal phosphopeptide EEEESSANGTASQstsPSQPR (1061-1083 region) on Ser-1074, Thr-1075 and Ser-1076. The phosphorylation status of these residues has not been investigated for their implication in CBP transactivation activity.

In addition to protein CREB1, the cAMP response element binding protein family also contains transcription factors CREM and ATF, which share a high degree of sequence and structural similarity. As described above, we observed differential abundance of the phosphopeptide containing Ser-271 on CREB1; interestingly, we also found the corresponding peptide of CREM harboring phosphosites Ser-271, Ser-274, Ser-277 and Ser-286 with 1.7-fold higher phosphorylation in high producer A1 cells. Despite being a close member of the CREB family and previously implicated in transcription regulation, CREM and ATF1 transcription factors have not been subjected to the same scientific inquiry as CREB1.

3.5 Concluding remarks

Transcription initiation occurs when the transcriptional machinery binds to the promoter regions of genes. Subsequently, mRNA synthesis occurs and then proteins are trafficked, folded and secreted. Recombinant gene expression in biopharmaceutical processes is frequently driven by the CMV promoter. Many studies have demonstrated the importance of employing vector design, UCOE, and MARs elements to influence the chromatin state to augment the expression of therapeutic proteins in mammalian cells (Veith, Ziehr, MacLeod, & Reamon-Buettner, 2016). However, there is little evidence of the role of the nuclear proteome and the interaction of DNA promoters with transcription factors to demonstrate the mechanism of transcriptional regulation. In this study, two cell clones exhibiting different productivity levels were investigated to determine the mechanism behind higher

productivity. From a previous characterization, higher productivity in the selected clones was determined to be influenced by the increased transcription within the higher producer cell lines.

Based upon the binding consensus of the transcription factors in the database, we hypothesized that CREB1 has a large likelihood of binding to the promoter region. By employing ChIP, we were able to study the interactions with the chromatin in each of the cell lines investigated. Our results indicate ~6-fold increased binding in the higher producer cell line. CREB1 is also implicated in other aspects of gene regulation, including histone modifications by CBP, a histone acetyltransferase, (Khan et al., 2017) (Levine et al., 2005) and in interactions with the DNA methyltransferases Dnmt3a and Dnmt3b (Hervouet, Vallette, & Cartron, 2009). As the consensus binding region of the transcriptional factors falls on the site of a CpG island within the CMV promoter, we expect a complex interplay in the function of transcription factor with the DNA in cells adapted to very high productivity phenotypes. Our understanding of the role of CHO nuclear proteome needs to evolve for us to understand the complex interplay between the transcriptional machinery and the high productivity phenotype.

Supplementary data

Histone deacetylases (HDACs) are a group of enzymes that remove acetyl groups from lysine residues in histones and non-histone proteins, resulting in transcriptional repression. HDACs have a role in cell growth arrest, differentiation and death. Methylation of the DNA and HDAC activity are closely linked.

We investigated the role of HDACs and related proteins in our cells using RNA extracted from exponentially growing cells. We observed no significant fold changes in the majority of the histone deacetylases or related proteins between in the cell lines. A detailed comparison of the transcriptomic (Fig), proteomic and phosphoproteomic changes in HDACs and associated proteins are given in Supplementary Table C3-2.

We investigated the role of HDACs and related proteins in our cells using RNA extracted from exponentially growing cells. We observed no significant fold changes in the majority of the histone deacetylases or related proteins between in the cell lines. A detailed comparison of the transcriptomic (Fig), proteomic and phosphoproteomic changes in HDACs and associated proteins are given in Supplementary Table C3-2.
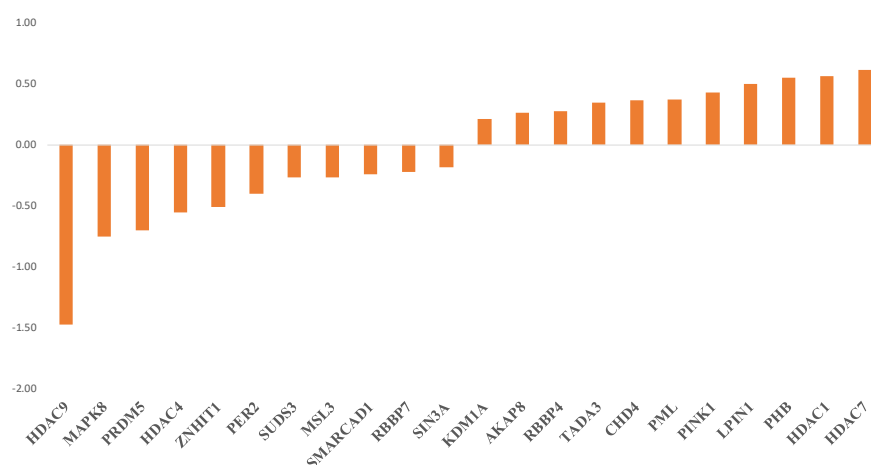
Supplementary Figure 3 Log$_2$ Fold-change (A1/A0) of mRNA levels for histone deacetylases and related proteins
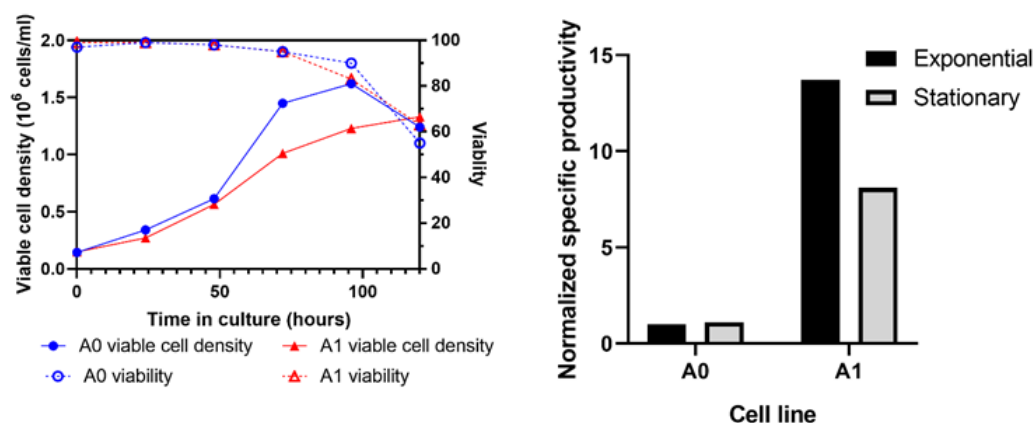
Supplementary Figure 4 Growth (left panel) and specific productivity (right panel) data for cultures used for RNA-seq and proteomic studies. Samples were taken at 48 hours (exponential) and 96 hours (stationary) for antibody titer measurements. Specific productivities are normalized to the specific productivity for A0 at 48 hours.

Supplementary tables can be downloaded from

https://drive.google.com/open?id=1XskMA0UiS2oHTOZkaGyAob65nbUihR_G

4. Chapter 4 Quantitative label-free LC-MS/MS comparative proteomic and phosphoproteomic analysis of CHO-K1 cells adapted to growth in glutamine-free media.

Abstract

This study aims to provide insights into the molecular mechanisms underlying adaptation of CHO-K1 cells to growth in glutamine-free media and potentially identifying critical signalling proteins and pathways that can serve as targets for rational cell line engineering and media optimisation. CHO-K1 cells were successfully adapted to glutamine-free conditions using a novel and simple one-step glutamine reduction strategy, which resulted in an adapted cell line with a comparable phenotype to the parental cells in terms of cell growth and viability. We performed a global quantitative proteomic and phosphoproteomic analysis of cells adapted to growth in glutamine-free media and compared them to parental cells grown in media containing 8 mM L-glutamine. Our results revealed that the adaptation process is accompanied by changes in proteins associated with cytoskeleton rearrangement and mRNA splicing as evidenced via functional analysis of 194 differentially expressed proteins between the two cell lines. Also, we identified 434 differentially expressed phosphoproteins, as a result of adaptation to L-glutamine-free conditions with an associated enrichment of pathways associated with MAPK and calcium signaling pathways. This work provides a comprehensive proteomic and phosphoproteomic analysis of protein expression changes after adaptation to glutamine-free growth conditions. We highlight critical pathways to consider when designing future studies to understand further and engineer glutamine metabolism, and rational design of improved feeding strategies. Data are available via ProteomeXchange with identifier PXD016692.

4.1 Introduction

Monoclonal antibodies (Mabs) continue to dominate biopharmaceutical product approvals. Up to 85% of new Mab products approved between 2014-18 were produced in CHO cells (Walsh, 2018). The last decade has seen remarkable improvements in the recombinant protein production capabilities of these cells. In recent years, CHO cell research has mainly focused on understanding the genomic, transcriptomic, and to some extent, the proteomics changes in these cells under different industrially relevant contexts (Yusufi et al., 2017). Other regulatory mechanisms, such as post-translational modification of proteins, e.g. phosphorylation, have been under-studied in CHO cells. Protein modification by phosphorylation can have a profound impact on the activity, conformation and localization of a phosphoprotein and with the ubiquitous presence of phosphorylation in the cellular proteome, virtually all regulatory mechanisms are directly or indirectly affected by phosphorylation events (Ramroop et al., 2018). A number of recent studies by our group and others have shown the additional benefit of including phosphoproteomic data in CHO cell studies to understand bioprocess-relevant phenotypes such as growth and productivity (Dahodwala et al., 2019; Henry et al., 2017; Kaushik et al., 2018; Schelletter et al., 2019). Very few studies have been carried out to investigate the proteomic and phosphoproteomic landscape as a result of nutrient deprivation, in particular of glutamine deprivation. Glutamine metabolism has been widely studied in cancer and in CHO cells, a system which is known to use this nitrogen and carbon source in a very efficient manner. Glutamine is metabolized to glutamate by the enzyme glutaminase and feeds into a large variety of metabolic pathways such as the TCA cycle, supporting the synthesis of ATP as a source of carbon, or the synthesis of amino acids, proteins or nucleotides as a source of nitrogen (Cluntun et al., 2017; Smith 1990).

In CHO cells, besides its role as a carbon, nitrogen and energy source, the interest in glutamine has mainly been focused on the challenges derived from its supplementation. It is by now well established that the metabolism of glutamine produces ammonia, a toxic byproduct for CHO cells with a detrimental effect on growth, productivity and, product quality ((Pereira et al., 2018). Attempts have been made to remove or substitute glutamine from CHO culture media with the aim to reduce toxic buildup and improve bioprocess output (Altamirano et al., 2000; Genzel et al., 2005; Ha and Lee 2014; Hong et al., 2010; Leong et al., 2018; Rajendra et al., 2012; Torres et al., 2019). As a radical solution, glutamine-free adapted cell lines have been established and characterized (Bort et al., 2010a; Taschwer et al., 2012).

In this study, we successfully adapted a CHO-K1 parental cell line to grow under glutamine free conditions using a one-step glutamine reduction approach. Following that, we performed proteomic and phosphoproteomic profiling of the obtained glutamine-free adapted cell line and compared it with the original parental cell line grown in glutamine supplemented media aiming to identify the critical differences between them.

4.2   Materials and Methods

4.2.1   Cell Culture

Parental CHO-K1 were routinely grown in BalanCD® media (Irvine Scientific) supplemented with 8 mM L-glutamine (Gibco) and 2% Polyvinyl-alcohol (PVA) as an anti-clumping agent. Glutamine-free adapted CHO-K1 cells were cultured in the same media but without glutamine supplementation. Cell cultures were routinely performed in Tubespin® bioreactor 50 tubes (TPP) (5 mL working volume) and in 250 mL shake flasks (20-50 mL working volume), and incubated in an orbital shaker (Kühner) at 37 ºC, 80% humidity and 5% CO_2 at a shaking speed of 170 RPM. Viable

cell density was monitored using the ViaCount™ assay on a GUAVA® EasyCyte™ benchtop flow cytometer (Merck Millipore, UK).

4.2.2 One Step Glutamine-reduction Adaptation Strategy

CHO-K1 cells were seeded at a standard cell density ($2x10^5$ cells/mL) in BalanCD® media with 2% PVA and were sub-cultured every 3-5 days in glutamine-free media for up to 17 passages, keeping them as much as possible in the exponential growth phase. These cells will be hereafter referred to as NN cells. The same original CHO-K1 population was maintained in BalanCD® media containing 8 mM L-glutamine and was sub-cultured and sampled at the same time points, and are hereafter referred to as GN cells. Before each passage, cell density and viability measurements were taken as described above.

4.2.3 Ammonia Quantification

Ammonia concentration measurements were performed on the culture supernatant using the L-Glutamine/Ammonia Assay Kit (Rapid) (Megazyme) in 96-well plate format and following the manufacturer's protocol and recommendations.

4.2.4 Growth Rate Calculations

Average growth rates were calculated for each passage from viable cell density (VCD) measurements taken before seeding the new tubes and taking into account that the cells were initially seeded at $2x10^5$ cells/mL, assuming exponential growth during that period. The formula used was μ [h^{-1}] = ln (VCD t_1 / VCD t_0) / t_1-t_0

4.2.5 Cell lysate preparation

Samples for cell lysates were obtained by spinning down $2x10^7$ cells for 5 minutes at $300{\times}g$, and the supernatant was discarded. Cell pellets were washed twice with ice-cold PBS and stored at -80°C until further processing. Cells pellets were resuspended in 8 M urea, 50 mM Tris, 75 mM NaCl (pH 8.2) buffer supplemented with 1X Halt

protease inhibitors (Thermo Fisher Scientific), 1X Halt phosphatase inhibitors (Thermo Fisher Scientific), and 0.1% ProteaseMAX™ surfactant detergent (Promega). Cell lysis was performed on ice using a handheld ultrasonicator at a medium setting by 30s sonication pulses for a total of three times per sample. After achieving a homogenous, free-flowing mixture, the cellular debris was removed by centrifugation of the cell lysate at 14 000×g for 10 min at 4 °C. The supernatant containing the protein fraction was transferred to a clean and pre-chilled microcentrifuge tube. Protein quantitation was performed by the Quick Start Bradford Protein Assay (BioRad).

4.2.6 In-Solution Protein Digestion and Phosphopeptide Enrichment

Proteomic samples equivalent to 1.2 mg of protein were taken from each sample for proteomic and phosphoproteomic analysis. Proteins were denatured by treatment with 0.5 M freshly prepared dithiothreitol (DTT) to a final concentration of 5 mM for 25 min at 56°C. Samples were then allowed to cool down to room temperature before being alkylated by the addition of a freshly prepared stock of 0.5 mM iodoacetamide to a final concentration of 14 mM at room temperature for 30 min in the dark. Samples were then diluted at a ratio of 1:5 in 25 mM Tris-HCl (pH 8.2), to reduce urea concentration from 8 M to 1.6 mM for peptide digestion using a double digestion approach with sequencing grade modified trypsin (Thermo Fisher Scientific) initially at 1:100 enzyme: protein ratio for 4 h at 37°C followed by the addition of the same amount of trypsin for incubation at 37°C overnight. After overnight digestion, any remaining trypsin activity was quenched by the addition of Trifluoroacetic acid (TFA) to a final concentration of 0.4%. Digested and acidified samples containing peptide mixture were cleaned up using a Sep-Pak C18 Vac cartridge (Waters) with negative pressure. The C18 cartridge was washed and conditioned by adding 9 mL of acetonitrile (ACN) followed by 3 mL of 50% ACN and 0.5% acetic acid. The cartridge

was then equilibrated with 9 mL of 0.1% TFA, and the samples were loaded in 0.4% TFA. Loaded samples were desalted with 9 mL 0.1% TFA. TFA was removed with 1 mL 0.5% acetic acid. Desalted peptides were eluted with 6 mL of 50% ACN, 0.5% acetic acid. At this stage, 10% of the peptide elution was aliquoted for complete proteome analysis and the remaining 90% of the sample was snap-frozen in liquid nitrogen for lyophilisation. Lyophilized peptides were phosphopeptide-enriched using Fe-NTA (IMAC) spin columns (Pierce, Thermo Fisher Scientific) as per manufacturer's instructions. The concentration of non-phospho-enriched and phospho-enriched peptides was quantified using a Nanodrop One spectrophotometer (Labtech International, UK).

4.2.7 LC-MS/MS Analysis

LC-MS/MS separations were performed using an UltiMate 3000 nanoRSLC system (Thermo Scientific) coupled in-line with an Orbitrap Fusion™ Tribrid™ Mass Spectrometer (Thermo Scientific) for MS acquisition, and essentially as previously described (Kaushik et al., 2018). For non-phospho-enriched samples, 1 µg peptides per sample were picked up by the autosampler and loaded onto the trapping column (PepMap100, C18, 300 µm × 5 mm) (Thermo Scientific) for 3 min at a flow rate of 25 µl min⁻¹. Peptides were then resolved on an analytical column (Easy-Spray C18 75 µm × 250 mm, 2 µm bead diameter column) (Thermo Scientific) using a gradient of 98% A (0.1% [v/v] formic acid (FA)): 2% B (80% [v/v] ACN, 0.08% [v/v] FA) to 35% B over 120 min at a flow rate of 300 nL min−1.

The Orbitrap Fusion™ Tribrid™ was set to acquire the MS1 spectra over m/z 200-1500 in the Orbitrap (120 K resolution at 200 m/z), and automatic gain control (AGC) was set to accumulate 4×10^5 ions with a maximum injection time of 100 ms (phospho: 50 ms). Data-dependent tandem MS analysis was performed using a top-

speed approach (cycle time of 3s). MS2 spectra were acquired in the ion trap. The intensity threshold for fragmentation was set to 5000 and included charge states 2+ to 7+ (phospho: included charge states 2+ to 6+). A dynamic exclusion of 50 s was applied with a mass tolerance of 10 ppm. AGC was set to 1×10^4 with a maximum injection time set at 35 ms. Additionally, for phospho-enriched samples, phosphosite identification was achieved by enabling multistage activation for neutral-loss triggered fragmentation for all precursor ions exhibiting neutral loss of mass 97.9763 with a mass tolerance of 0.5 m/z, where the neutral loss ion was one of the top 10 most intense MS2 ions. AGC was set to 20,000 with a maximum injection time set at 90 ms. The mass spectrometry proteomics and phosphoproteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al., 2019) with the dataset identifier PXD016692.

4.2.8 Quantitative Label-Free LC-MS/MS Analysis

Differential expression values for non-phospho-enriched and phospho-enriched peptides were obtained in separate experiments via Progenesis QI for Proteomics (Nonlinear Dynamics). The LC-MS/MS files were uploaded to Progenesis QI for Proteomics for automatic alignment of samples runs to account for run to run variability between each sample. Aligned samples were then grouped as GN and NN samples and the quantified ion abundances for peptide ion peaks were filtered based on an ANOVA p-value <0.05 between experimental groups. The output from Progenesis QI containing spectral information for differential identified features in a mascot generic file (mgf) was searched for protein identification using SequestHT search algorithm in Proteome Discoverer 2.2 (Thermo Fisher Scientific). The peak list was searched against a recently published CHO proteogenomic database (Li et al., 2019). For phosphosite identification, the PhosphoRS algorithm (Taus et al., 2011)

was run through ProteomeDiscover 2.2 while searching in SEQUEST and by applying a phosphosite probability score of 75% or higher for S, T, or Y amino acids in PhosphoRS (Henry et al., 2017). Data were filtered to a 1% false discovery rate (FDR) on PSMs.

4.2.9   Gene Ontology Analysis

Gene symbols for identified differentially expressed proteins and phosphoproteins were imported into DAVID (https://david.ncifcrf.gov) (Huang et al., 2009), ToppGene Suite (https://toppgene.cchmc.org) (Chen et al., 2009) online gene ontology platforms for functional pathway analysis. To refine the number of pathways identified, an adjusted p-value of ≤0.05 between experimental groups was used. For each phosphoprotein of interest, the phosphosite number was assigned from the corresponding human protein sequence match. The functional information of such phosphosites was obtained from phosphosite.org (www.phosphosite.org), an online resource for providing information and tools for the study of protein post-translational modifications. The protein-protein interaction networks were generated via STRING Database (Szklarczyk et al., 2019) and visualized in the Cytoscape app (Lopes et al., 2011). Other data visuals were created through the ggplot2 package (Wickham, 2009) in the R programming environment for statistical computing (R Core Team, 2014). Overlapping identifications are shown as Venn diagrams created from Venny 2.1 (Oliveros, 2007).

4.3   Results and Discussion

4.3.1   A one-step glutamine reduction approach generates glutamine-free adapted CHO cell line with minimal impact on cell viability.

In order to assess the impact on cell growth and viability of a sudden reduction of L-glutamine in the media, CHO-K1 cells were seeded in 0 to 8 mM L-glutamine

containing medium, and cell density and viability were monitored over time (Supplementary Figure 5). Cells growing in media containing up to 2 mM of L-glutamine did not show differences in growth during the first 72 hours of culture and then reached higher cell densities proportional to initial L-glutamine concentration. Cells seeded in media without L-glutamine exhibited a slower growth profile and reached lower cell densities. Interestingly, cell viability remained above 90% for all conditions until the late stages of the culture, even in the tubes without L-glutamine. Based on this observation, a single step for glutamine reduction strategy was undertaken. This involved the complete removal of L-glutamine from the culture media (no supplementation) followed by sub-culturing of the cells for an extended period of time Figure 4.1. A shows the calculated average growth rate for each passage and the corresponding viability measurements for the CHO-K1 cells maintained in glutamine-free media and the CHO-K1 cells maintained in glutamine-containing media over the course of the 17 passages. CHO-K1 cells grown in the absence of L-glutamine exhibited lower average growth rates during the first 3 passages. Interestingly, from passage 4, a progressive increase of the average growth rates for glutamine-free cells up to within a 10% range of the growth rates was observed for CHO-K1 cells growing in 8 mM L-glutamine over the same period of time. After 17 passages, no changes in the average growth rates were observed, and cells were considered to be adapted. In terms of viability, cells exhibited a drop of 20% during the first two passage and progressively recovered viability levels to over 95% from passage 4 onwards. In order to compare adapted and parental CHO-K1 cells, a growth curve experiment was set to compared growth profile, maximum cell density, viability and ammonia concentration in the media for both cell lines (Figure 4.1B). CHO-K1 cells adapted to glutamine-free conditions (NN) exhibited slightly lower specific

growth rates than parental CHO-K1 cells growing in glutamine containing media (Supplementary Figure C4-2), and reached higher cell densities than the original CHO-K1 cell line growing in glutamine containing media. In addition, the ammonia concentration in the media was lower for the glutamine-free adapted CHO-K1 cell line and remained constant at around 0.05 g/L for the whole duration of the culture, in comparison to the CHO-K1 cell line growing in L-glutamine containing media, which showed increasing ammonia concentrations over the time ranging from 0.1 g/L after 48h to >0.25 g/L at the end of the culture, 192h (Figure 4.1B). The adaptation of CHO cells to glutamine-free media is not new as it has previously been successfully achieved in the past using a FACS-directed evolution strategy (Bort et al., 2010b). However, to our knowledge, the results are the first evidence that CHO cells can adapt to glutamine-free conditions without any additional intervention. Also, these results challenge previous reports as no lethal effect from a sudden glutamine reduction other than an early drop in viability was observed. Bort et al., (2010b) report an initial drop in growth rate and viability after reducing glutamine content in the culture media, which is consistent with our findings. However, they also report a decrease in viability after the routine passage of cells in medium with reduced glutamine content. It must be noted that in terms of time, our one-step adaptation approach was similar to the FACS-based methodology published in the same study, with a total time of approximately ten weeks.
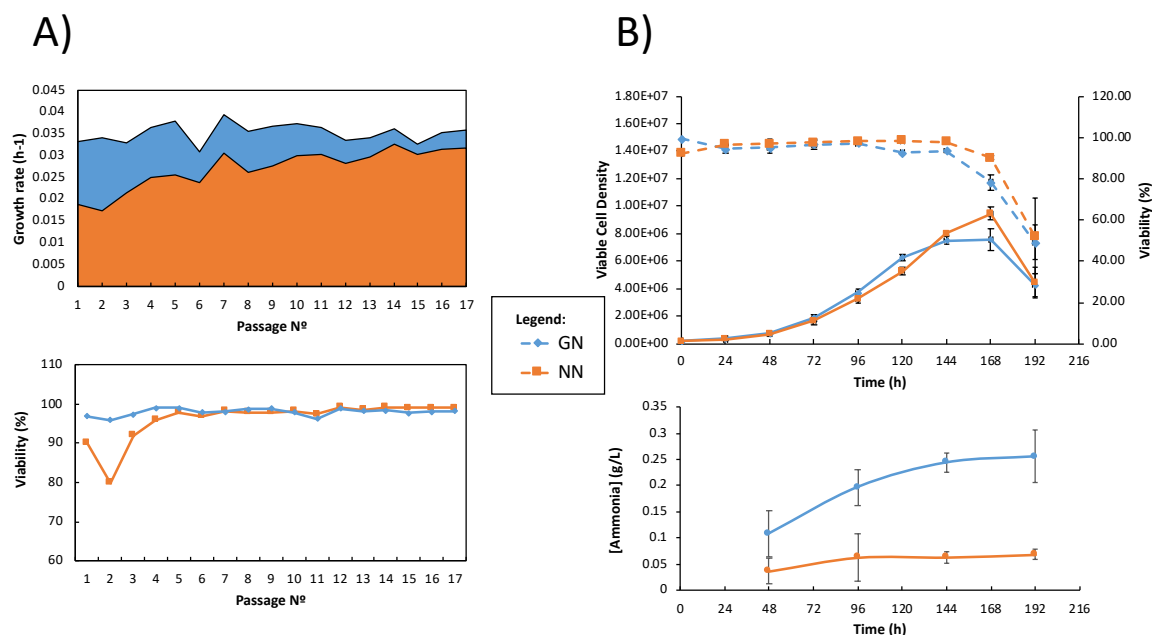
Figure 4.1 Adaptation of CHO-K1 cells to L-glutamine free media. (A) growth rates (top) and viability (bottom) over the course of the one-step glutamine-free adaptation process. CHO-K1 cells were sub-cultured in glutamine-free media until adaptation was completed (NN - orange). CHO-K1 cells were sub-cultured in media containing 8 mM L-glutamine as the control (GN - blue). (B) Final growth curves of the glutamine-free adapted CHO-K1 cell line (orange) compared with the glutamine dependent parental cell line (blue): viable cell density and viability (top) and ammonia concentration in the media (bottom).

4.3.2 Proteomic and phosphoproteomic changes related to glutamine-free adaptation

For the identification of proteomic and phosphoproteomic differences between GN and NN cells, three independent biological replicates per condition were prepared for LC-MS/MS analysis. Prior to the differential proteomic analysis, each sample run was individually searched against the CHO proteogenomic database using the SequestHT search algorithm in Proteome Discoverer 2.2. A detailed list of all identifications is provided in Supplementary Table C4-1. In summary, the number of identifications was consistant among sample type (see Supplementary Figure 7), i.e., from the non-phospho-enriched samples, an average of 28117 peptides per sample (CV= 1.7%) and an average of 3347 total proteins (unique peptides $\geqslant$ 2) per sample was identified (CV=1.50%). Similarly, from the phospho-enriched fractions, an average of 6126 phosphopeptides with the PhosphoRS phosphosite probability of greater than 75% per

sample was identified (CV=3.37%) corresponding to the identification of an average of 2177 phosphoproteins per sample (CV= 2.20%). Overlaping identifications among biological replicates of each condition, i.e., GN and NN, are shown in Supplementary Figure 8. Three biological replicates from the non-phospho-enriched samples of the GN sample group showed a 73.8% overlap in protein identifications, and three replicates of the NN sample group showed 73% overlapping protein identifications. The phosphoproteins identified from three biological replicates of GN and NN samples showed an overlap of 53% and 57%, respectively. The overlap of 73% in protein identifications and a 53% to 57% overlap in phosphoprotein identifications among the biological replicates is on par with previous studies (Hinzke et al., 2019) (Murillo et al., 2018). A lower overlap of phosphoproteins could be due to the labile nature of phosphorylation modifications (Hauser et al., 2017). The combined search of all non-phospho-enriched samples identified 38100 peptides corresponding to 4049 proteins with unique peptides ≥ 2 and the combined search for all phospho-enriched samples identified 8271 phosphopeptides corresponding to 2174 phosphoproteins. Furthermore, in agreement with previous studies that use IMAC chromatography for phosphopeptide enrichment (Thingholm et al., 2009), our analysis also reports the enrichment of predominantly singly phosphorylated peptides. As shown in Figure 4.2A, 80% of all phosphopeptides were phosphorylated at one site, whereas 16% of phosphopeptides were doubly phosphorylated, and 4% of the phosphopeptides were triply phosphorylated. In total, 9947 phosphosites were identified with a PhosphoRS best site probability greater than 75%. The distribution of these phosphosites, according to phosphorylated residue, is shown in Figure 4.2B where 8899 (88%) were serine phosphorylated, 1068 (11%) were threonine phosphorylated, and 80 (1%) of the identified phosphosites were phosphorylated at a tyrosine residue.
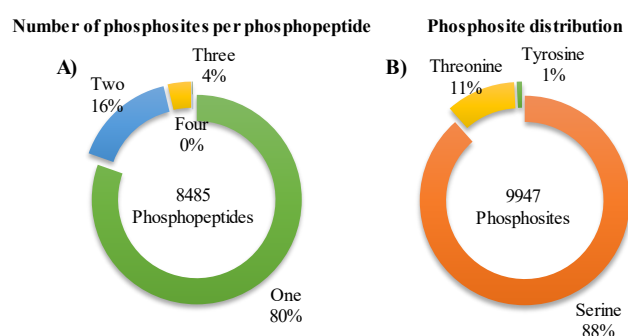
Figure 4.2 A) Number of phosphosite per phosphopeptide identified from the combined analysis of phospho-enriched samples. B) Distribution of all phosphosites according to the amino acid residue of phosphorylation.

Quantitative total proteomic analysis of non-phospho-enriched samples from GN cells and NN cells revealed changes in the expression levels of 194 proteins. From the phosphoproteomic analysis, 434 phosphoproteins were found to be differentially expressed between the GN and NN cells with a fold change $\geq$ 1.5. 1494 phosphopeptides were found to be differentially expressed between the two cell lines; with some phosphoproteins contained multiple phosphosites. Figure 4.3A and Figure 4.3B shows the Pearson correlation matrix of all such protein and phosphoprotein identifications from the proteomic and phosphoproteomic analysis, respectively. Non-supervised sample clustering indicates the robustness of the data collected. Moreover, Figure 4.4A and Figure 4.4B show fold change distribution of all confidently identified (p-value<0.05) proteins and phosphoproteins identified by Progenesis QI for Proteomics in a Volcano plot where the differentially expressed proteins and phosphoproteins are shown in red, and the top 20 most differentially expressed proteins and phosphoproteins based on fold-change are highlighted. A complete list of all differentially expressed proteins, phosphoproteins and phosphopeptides is shown in Supplementary Table C4-2. An numerical overview of all differential proteins, phosphoproteins and phosphopeptides identified between NN and GN cells is shown in Table 4.1.
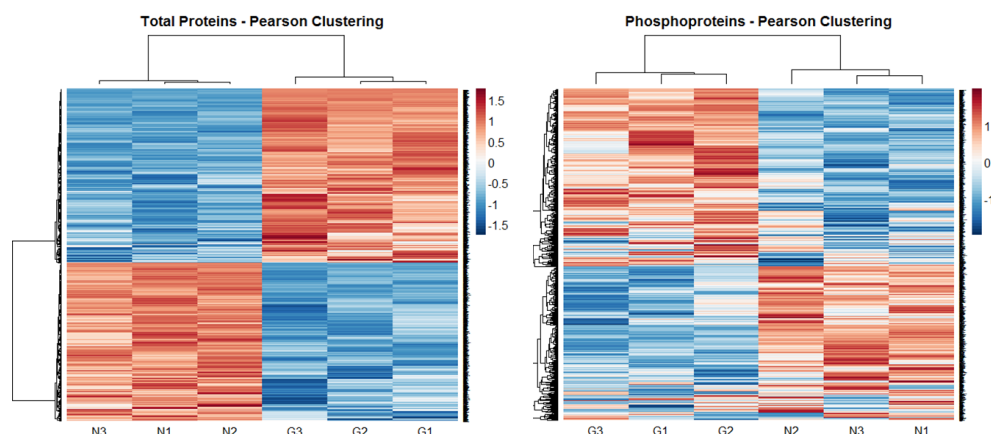
Figure 4.3 - Unsupervised Pearson clustering of all differentially expressed proteins (A) and phosphoproteins (B) shows sample clustering into distinct groups that are NN and GN cells. Sample replicates are shown (labelled 1-3) for each sample type.

Figure 4.4 - Volcano plots of proteins (left) and phosphoproteins (right) identified by label-free LC-MS/MS comparative analysis using Progenesis QI for Proteomics. Each point represents the fold change between two sample groups plotted against the level of statistical significance. Differentially expressed proteins and phosphoproteins with fold change >1.5 are shown in red.

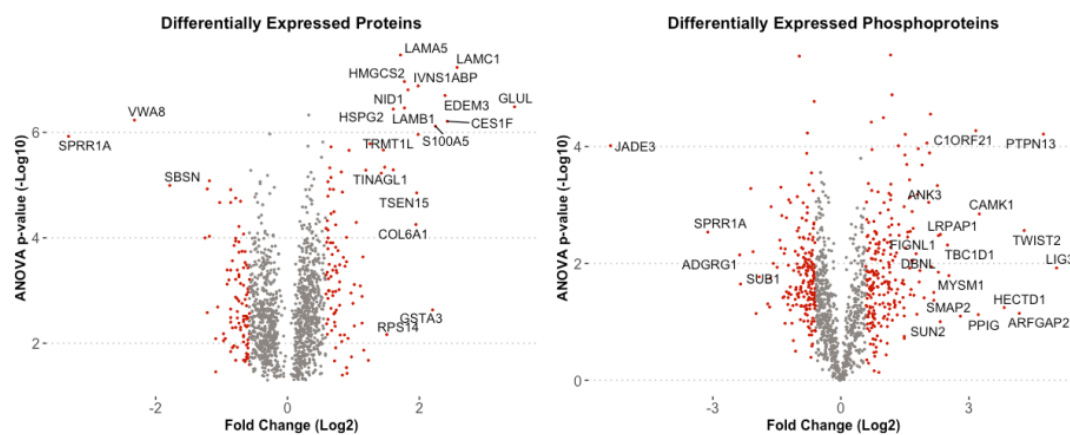*Table 4.1 - Overview of total differential proteins, phosphoproteins and phosphopeptides identified between NN and GN cells. Numerical values are of identifications for each category relative to expression in NN cells.*

	DE Proteins	DE phosphoproteins	DE Phosphopeptides
High in NN	100	233	821
Low in NN	94	201	673
Total	194	434	1494

It should be noted that glutamine synthetase (GLUL) was the most differentially expressed protein with a 10.9 fold increased expression in the NN cells. Glutamine synthetase catalyzes ATP-dependent conversion of glutamate and ammonia to glutamine. As a result of glutamine starvation, cells expressed a higher levels of glutamine synthetase. Studies have previously shown the increased activity of glutamine synthetase in CHO cells in the absence of glutamine (Zhang et al., 2006). Here, we show that this increased activity could be due to the increased expression of protein GLUL in agreement with previous cancer studies that have shown increased GLUL expression under glutamine limitation conditions (Kung et al., 2011). The second-most differentially expressed protein was Coniferin-A (SPRR1A). In the absence of glutamine, the total protein expression levels of SPRR1A were reduced by 9.95-fold in NN cells. This protein belongs to the small proline-rich protein family, and has been characterized as a structural protein that becomes cross-linked with other membrane proteins by transglutaminase during the differentiation of squamous epithelial cells (Steinert & Marekov, 1997). To our knowledge, there are no studies relating its function with a nutrient deprivation scenario.. However, the protein contains a high number of glutamine residues (Jing et al., 2012) which could explain

its low expression levels in the absence of extracellular glutamine. Figure 4.5 shows the Progenesis QI for Proteomics output for all peptides associated with GLUL and SPRR1A.
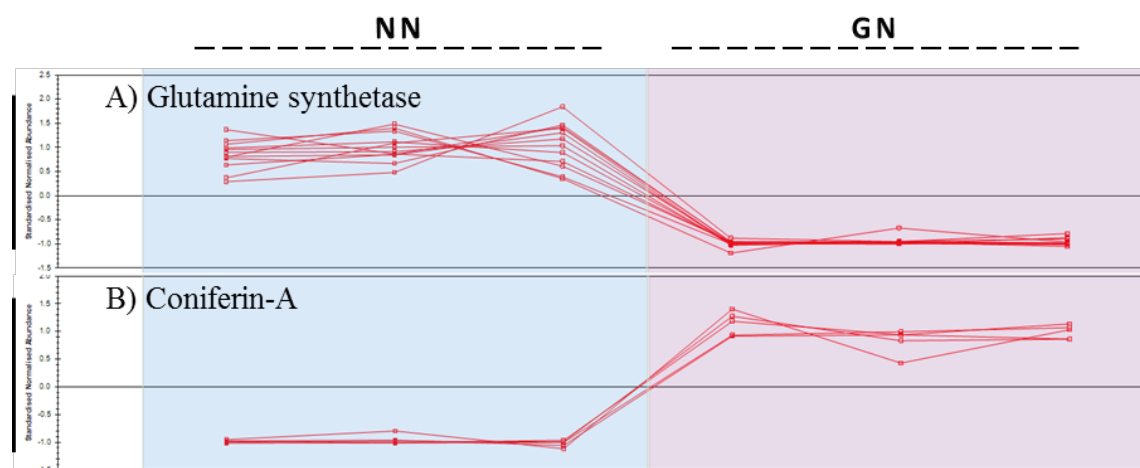
Figure 4.5 Total protein expression change of protein glutamine synthetase (Top) and Coniferin-A (Bottom). Progenesis QI output showing standardized expression change for each contributing peptide. Three replicates per condition are shown on the x-axis whereas y-axis represents standardized expression change for each peptide for the protein.

4.3.3 Pathway analysis of differentially expressed proteins.

All differentially expressed proteins and phosphoproteins were placed into broad functional enrichment categories based on the output from the Toppgene functional enrichment tool. The top significant enriched terms from GO: Biological function and GO: Pathway analysis are shown in Figure 4.6. Dominating categories for GO biological process were mainly RNA processing and extracellular matrix (ECM) organization. Splicing of pre-RNA to mature mRNA is a critical step in the regulation and control of gene expression (Oltean & Bates, 2014), and alternative splicing produces different isoforms of the same gene (Black, 2003); hence, differential expression of RNA splicing factors can have a profound impact on cell biology (Koedoot et al., 2019). Despite the evidence of improved therapeutic protein yield through optimizing transgene splicing in CHO cells (Lucas et al., 1996), the understanding of systemic CHO splicing machinery is lacking. Our results show the

differential expression of 19 proteins associated with RNA splicing, pointing toward an important role of this regulatory mechanism in adaptation to glutamine-free media. The GO: biological process category enrichment of proteins associated with ECM structure organization such as three proteins from the Laminin family (LAMC1, LAMA5, LAMB1), two Laminin interacting proteins (NID1 and, HSPG2) and one Integrin protein (ITGB5) was also observed. Differential expression of proteins associated with cytoskeleton rearrangement was also seen in GO: Pathway analysis category, with terms such as regulation of cytoskeleton and vesicle transport among the top 10 enriched terms. The ability of CHO cells to maintain structural integrity in the bioreactor when scaling up to 20,000 litres is one of the reasons for their dominance in the industrial production of biopharmaceutical (Hu et al., 2011). In addition, growing evidence supports the connection of the cytoskeleton to autophagy, one of the central self-degradation cellular processes up-regulated upon starvation. For example, cytoskeleton rearrangement has been showed to participate in the early events of autophagy (Aguilera et al., 2012) and cytoskeletal proteins have been shown to play a crucial role in all its phases (Kast & Dominguez, 2017). In line with these findings, a recent transcriptomic study of a feed spike in an IgG producing CHO cell line also identified differential expression of various genes associated with the cytoskeleton component between feed-spiked cultures versus non-spiked during the late stages of the culture (Reinhart et al., 2018). Our results further demonstrate that a change in media composition can have a profound impact on the proteomic landscape of CHO cells with potential to change bioprocess related phenotypic traits such as changes in spliceosome machinery that regulates the conversion of pre-mRNA to mature RNA, a crucial step in the regulation of gene expression including the expression of a gene of interest. Furthermore, the cellular cytoskeleton does not only

provide structural integrity but is also actively involved in the regulation of cellular signaling processes (Forgacs et al., 2004), hence changes in the expression of structural proteins not only changes cellular morphology but it can also bring about changes in bioprocess relevant signaling pathways such as autophagy, translation (S. Kim & Coulombe, 2010), and protein synthesis (Gross & Kinzy, 2007).
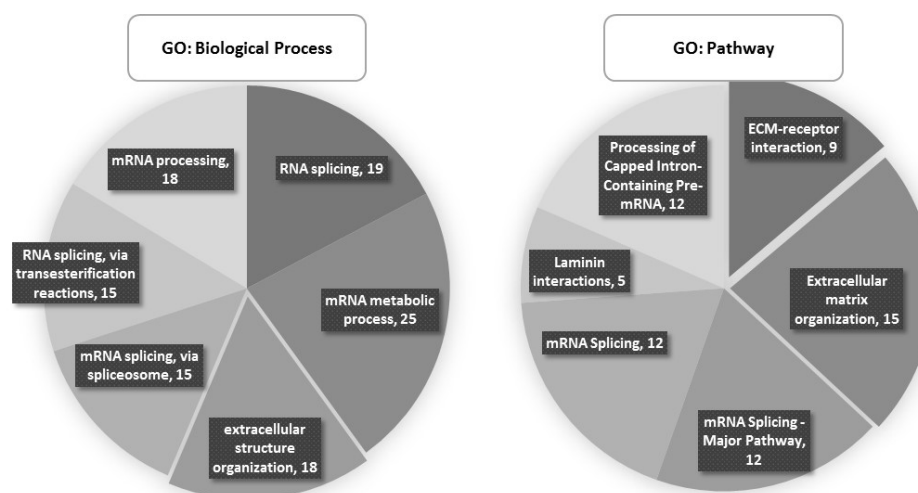
Figure 4.6 - Pie chart of gene ontology biological processes (left) and pathway enrichment (right) of differentially expressed proteins between GN and NN cells grouped from Toppgene functional enrichment output. The number of proteins and associated enrichment terms are displayed in the corresponding sections of the pie-chart.

4.3.4 Pathway analysis of differentially expressed phosphoproteins

The functional enrichment analysis of 434 differentially expressed phosphoproteins was performed via the Toppgene gene enrichment suite, and the top five most significant terms for each Gene Ontology categories are shown in Figure 4.7. Briefly, in terms of enriched biological processes (GO: Biological process) cytoskeletal protein binding with 60 differential phosphoproteins identified and GTPase regulator activity with 26 differential phosphoproteins identified showed the most significant enrichments. GO: Molecular function analysis also revealed an enrichment of proteins involved in the regulation of organelle organization function, protein localization and cytoskeleton rearrangement. The cellular component analysis also showed a

significant enrichment of proteins (n=69) from the cell junction component. Once more, these findings, along with the other enriched gene ontology terms shown in Supplementary Table C4-3, suggest widespread changes in cytoskeleton organization in response to long-term glutamine starvation. This observation is further strengthened by the results obtained from the pathway enrichment analysis performed using the KEGG and Reactome pathway databases (see Supplementary Table C4-3). The top enriched pathway from the KEGG database was the regulation of actin cytoskeleton. The output of enriched pathways from the Reactome database also highlighted the signalling of Rho-GTPases and signalling by the Robo receptor. Both pathways have been linked to the maintenance of cellular cytoskeleton, especially in cancer studies (Gara et al., 2015; Svensmark and Brakebusch 2019). In addition, RhoA has been shown to activate autophagy under nutrient deprivation conditions (Aguilera et al., 2012).

To gain deeper insights into the role of differentially expressed phosphoproteins in the NN cell phenotype all differentially expressed phosphoproteins were divided into two subgroups (Figure 4.8) based on their direction of expression change, i.e., subset 1- 233 phosphoproteins that were identified to be more phosphorylated in NN cells and subset 2- 201 proteins that were identified as having low phosphorylation in NN cells compared to GN cells. Both groups were subjected to protein-protein interaction analysis using STRINGdb database (Szklarczyk et al., 2019). K-means clustering within these networks was employed to identify network clusters and key proteins with high connections in the phosphoproteomic data. Two distinct clusters per subset were identified and are shown in Figure 4.8. Cluster 1 contains 26 proteins from which 7 proteins were from the MAPK signaling pathway. Central to this pathway are MAPK1 (ERK2), and MAPK3 (ERK1) with both proteins known to play a critical

role in the regulation of cell growth and differentiation (Sun et al., 2015). The phosphopeptide VADPDHDHTGFLT*EY*VAT*R (Thr-183, Tyr-185 and, Thr-188) from MAPK1 was found to be ~2 fold increased in the glutamine starved NN cells. This particular phosphopeptide represents a cluster of phosphorylation sites referred to as the ERK2 activation loop (Pegram et al., 2019). Dual phosphorylation of Thr-183 and Tyr-185 by the upstream kinase MEK1-2 is crucial for ERK2 activation (Aoki et al., 2011). Phosphopeptide IADPEHDHTGFLT*EYVAT*R (Thr-183, Thr-188) of ERK1 activation pool showed a 1.8 fold increased expression in NN cells suggesting the activation of ERK1/2 under glutamine-free conditions. Upon activation, MAP kinase phosphorylates cytosolic phospholipase A2 (cPLA2) on Ser-505 (Hirasawa et al., 1995). We found a 2.1 fold increase in expression of the phosphopeptide VHNFMLGLNLNTSYPLS*PLR (Ser-505) from cPLA2 in glutamine starved cells. Recently cPL2 activation has been linked to disruption of autophagy flux and neuronal cell death after spinal cord injury (Y. Li et al., 2019). Activation of the MAPK signaling cascade also promotes activation of NF-kB protein, an important group transcriptional factor that plays a critical role in the regulation of various key cellular process such as stress-regulated cell survival, plasticity and apoptosis (Mattson and Meffert 2006; Mincheva et al., 2011). In the context of recombinant protein production from CHO cells, CRE-binding transcription factors have been found to regulate the CMV promotor associated transgene expression by an increase in NF-kB1 (Brown et al., 2015). A recent CHO phosphoproteomic study that showed increased association on CREB1 protein with CMV promoter in high producer CHO cell line also predicted increased NF-kB1 activity (Dahodwala et al., 2019). Here we found a 1.87 fold increased expression of the phosphopeptide KLS*FTESLPGDSSLLSLNK (Ser-937) from NF-kB1 in NN cells while also reporting a 1.53 fold higher expression of NF-

kB2 total protein. Another member of the NF-kB family, the transcription factor RelB, was differentially phosphorylated between the analyzed conditions. A 2.2 fold increased expression of the RelB phosphopeptide LVS*PGPGPRPHLVITEQPK (Ser-96_Mouse) was observed in the NN cells. Unlike other members of NFkB family of proteins RelB does not directly bind to the DNA response element sequence but plays a crucial role in the regulation of apoptosis (Ge et al., 2016; Tan et al., 2015). Activated ERK1/2 also phosphorylates ERF, a potent transcription repressor (Sgouras et al., 1995); phosphorylation of Thr-526 on ERF is crucial for its translocation from the nucleus to the cytoplasm thus inducing transcription via the ETS promotor (Le Gallic et al., 2004). We identified increased expression of an ERF phosphopeptide GDTAAAAGPGEAGGPLT*PR (Thr-526) by 2.3 fold in NN cells. Four other members of the MAPK1 cascade were shared with cluster 3 of the phosphoproteins identified with decreased expression in NN cells. For example, phosphorylation of SRC, a non-receptor tyrosine kinase that plays a key role in cell-adhesion migration and stability (Frame, 2004) at Ser-17 at S*LEPAENVHGAGGAFPASQTPSKPASADGHR was decreased two-fold in NN cells. Ser-17 of SRC is phosphorylated by PAK1 which in turn is activated by protein CK2 by phosphorylation at Ser-223 and Ser-144 (Shin et al., 2013). From CK2 we found the phosphopeptide YMS*FTDK (Ser-144) to be 2.5 fold higher expressed in NN cells and DVATS*PIS*PT*ENNTAPPDALTR (Ser-220, Ser-223 and Ser-225) to be 2.14 fold higher phosphorylated in NN cells.
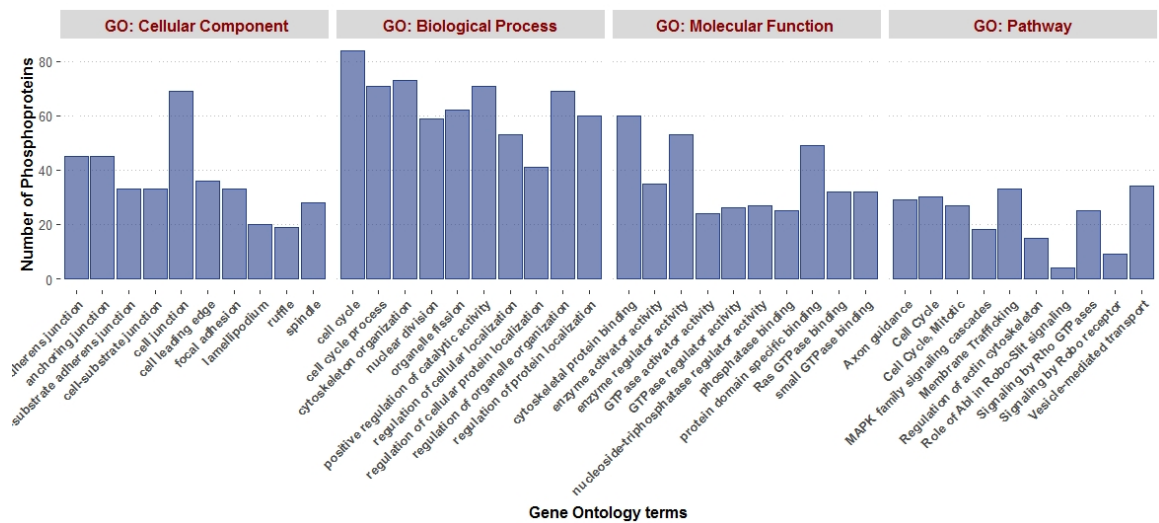
Figure 4.7 Gene Ontology analysis of differentially expressed phosphoproteins. The top 10 most significantly (p-value <0.05) enriched terms per gene ontology category are shown in a bar chart where the bar height is representative of the number of proteins in the GO term.
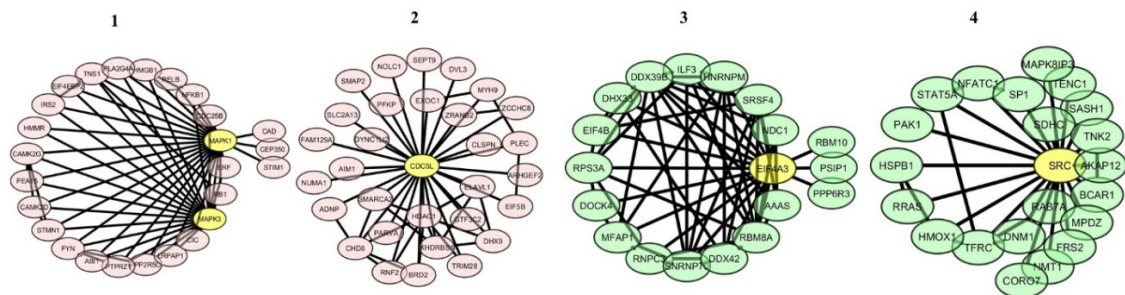
Figure 4.8 K-mean clustering of differentially expressed phosphoproteins. Large degree clusters identified from differentially expressed proteins that were identified to be high phosphorylated in NN cells (pink) and proteins that were identified to be low phosphorylated in NN cells (green).

4.3.5 Overlap between proteomic and phosphoproteomic differentially expressed proteins

To further gain insights into phosphorylation events underlying the adaptation to glutamine-free conditions, the overlap between proteins identified from the total proteome analysis and proteins identified from the phosphopeptide enrichments was investigated (see Supplementary Table C4-2). Figure 4.9A shows a Venn diagram of overlapping identifications (Oliveros, 2007) with 31 differentially expressed proteins found to be present in both the proteomic and phosphoproteomic lists. The expression profiles of these common proteins and phosphoproteins are shown in Figure 4.9B. The

breakdown of this overlap is as follows; sixteen proteins showed increased expression in the NN cell line at both the total protein and the phosphoprotein level. In this study, we also report increased expression of the NDRG1 phosphopeptide, T*AS*GS*SVTSLEGPR, which was found to be phosphorylated on Thr-328, Ser-330 and, Ser332 in the NN cell line and a 1.76 fold increased expression of total NDRG1 proteins in the NN cells. Previous studies have shown increased expression of phosphopeptide T*AS*GS*SVTSLEGPR of NDRG1 protein at cells growing at low temperature (Henry et al., 2017). Fourteen other proteins showed a decrease in expression in both datasets in the glutamine-free-adapted NN cells. From this analysis, two proteins show striking differences in terms of the changes in the total proteome fraction and the phospho-enriched fraction: CAMK2G and CAMK1. Calcium/calmodulin-dependent protein kinase type II subunit gamma (CAMK2G) was found to show increased phosphorylation despite a 1.6 fold decreased expression of total CAMKG2 in GN cells. Two CAMK2G phosphopeptides were found to show increased expression; the phosphopeptide QET*VECLR (Thr-287) was 1.9 fold increased in NN cells, and the phosphopeptide SDGGVKEPQTTVVHNAT*DGIKGS*T*ES*CNTTTEDEDLK (Thr-375, Ser-381, Thr-382 and, Ser-384) was 2.9 increased in NN cells. Ca^{2+}/calmodulin-dependent protein kinase such as CAMK1 and CAMK2, are central players in calcium-mediated activation of the MAPK pathway and have been shown to be involved in autophagy regulation as a response to amino acid starvation (Zhou et al., 2015). There are several reports providing evidence of autophagy occurring in CHO cells as a result of nutrient deprivation (Han et al., 2011;Sun and Lee 2008), and particularly after glutamine deprivation (Jardon et al., 2012). Management of autophagy has been suggested as a novel way to improve cell phenotypes for the production of biopharmaceuticals (Kim

et al., 2013). Our results suggest that further investigation of Ca^{2+} signalling and MAPK pathways could lead to novel strategies to manipulate these pathways in to further improve and control CHO cell phenotypes.

Cellular activity through phosphorylation is maintained by an intricate system of kinases that phosphorylate a phosphoprotein and phosphatases that remove the phospho-modification, and both processes are capable of changing protein activity such as migration, activation, repression, etc. Therefore, it is essential to study both phosphorylated and dephosphorylated protein levels in comparative phosphoproteomics. For example, thirty-nine phosphoproteins showed higher expression in the NN cells whilst the total protein expression was not significantly changed, including proteins MAPK1 as discussed previously. In contrast, 43 phosphoproteins showed lower expression in NN cells, but total protein expression did not change significantly, including four proteins from the mitophagy pathway (SRC, RAB7A, TBC1D15, RRAS2).
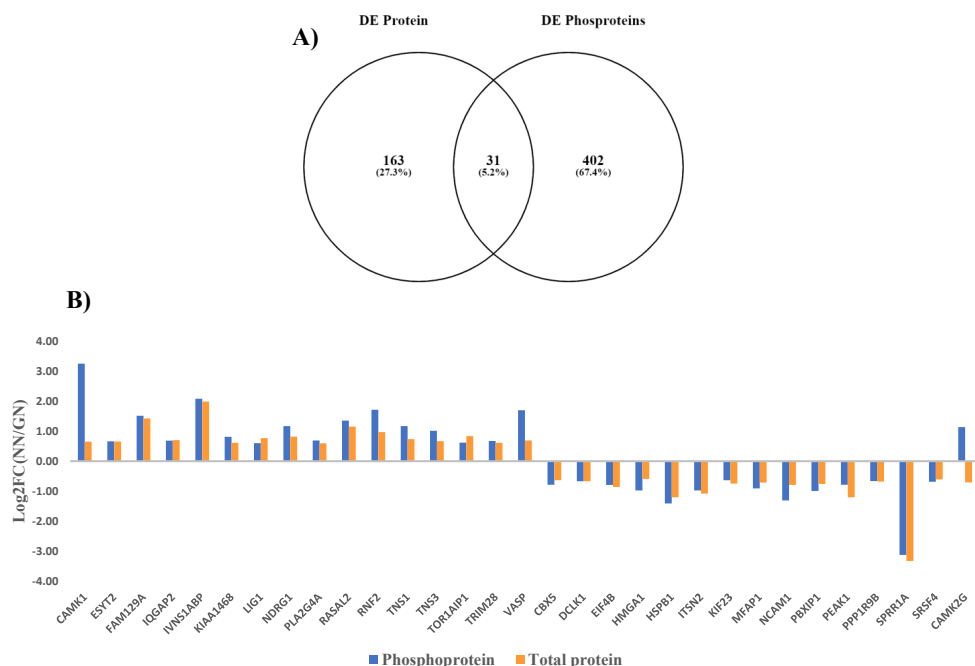
Figure 4.9 Overview of overlap between proteomic and phosphoproteomic differentially expressed proteins. (A) Venn diagram of differentially expressed proteins and phosphoproteins. (B) Bar chart of 31 proteins that were differentially expressed in both proteomic and phosphoproteomic analysis. The bars represent the magnitude and direction of fold change.

4.4 Conclusions

Glutamine removal and supplementation have been an active area in CHO cell research; previous studies have focused on the transcriptomic and metabolic changes associated with glutamine removal. In this study, we successfully adapted CHO-K1 cells to growth in glutamine-free media using a straightforward one-step glutamine reduction strategy, and we provide the first account of global proteomic and phosphoproteomic differences resulting from this adaptation. Our proteomic analysis showed differential expression of proteins associated with mRNA processing via the spliceosome, and cytoskeleton rearrangement. Both processes are highly relevant in CHO cells bioprocessing. Pre-mRNA splicing to form mature RNA is a critical step in the regulation of gene expression, and an increased understanding of the splicing machinery could provide an opportunity to fine-tune the expression of specific genes

136

towards recombinant protein production (Fallot et al., 2009). Both the proteomic and phosphoproteomic data suggest rearrangement of the cytoskeleton and disruption in cellular signaling pathways. Our phosphoproteomic analysis highlighted changes related to pathways such as ERK1/2, MAPK and Ca^{2+} signaling pathways as a result of cellular adaptation to glutamine starvation. These pathways have roles in the activation of autophagy, which could be postulated as an underlying cell survival strategy that allows glutamine-free adapted cells to thrive in the absence of glutamine supplementation. These findings could be extrapolated to suggest that similar changes could be occurring in the situation of nutrient deprivation such as those occurring in late phases of production batch cultures. In summary, this proteomic and phosphoproteomic study does not only expand our knowledge on the proteomics changes related to glutamine metabolism but has also provided useful information to be used towards developing novel CHO cell engineering efforts in the future.
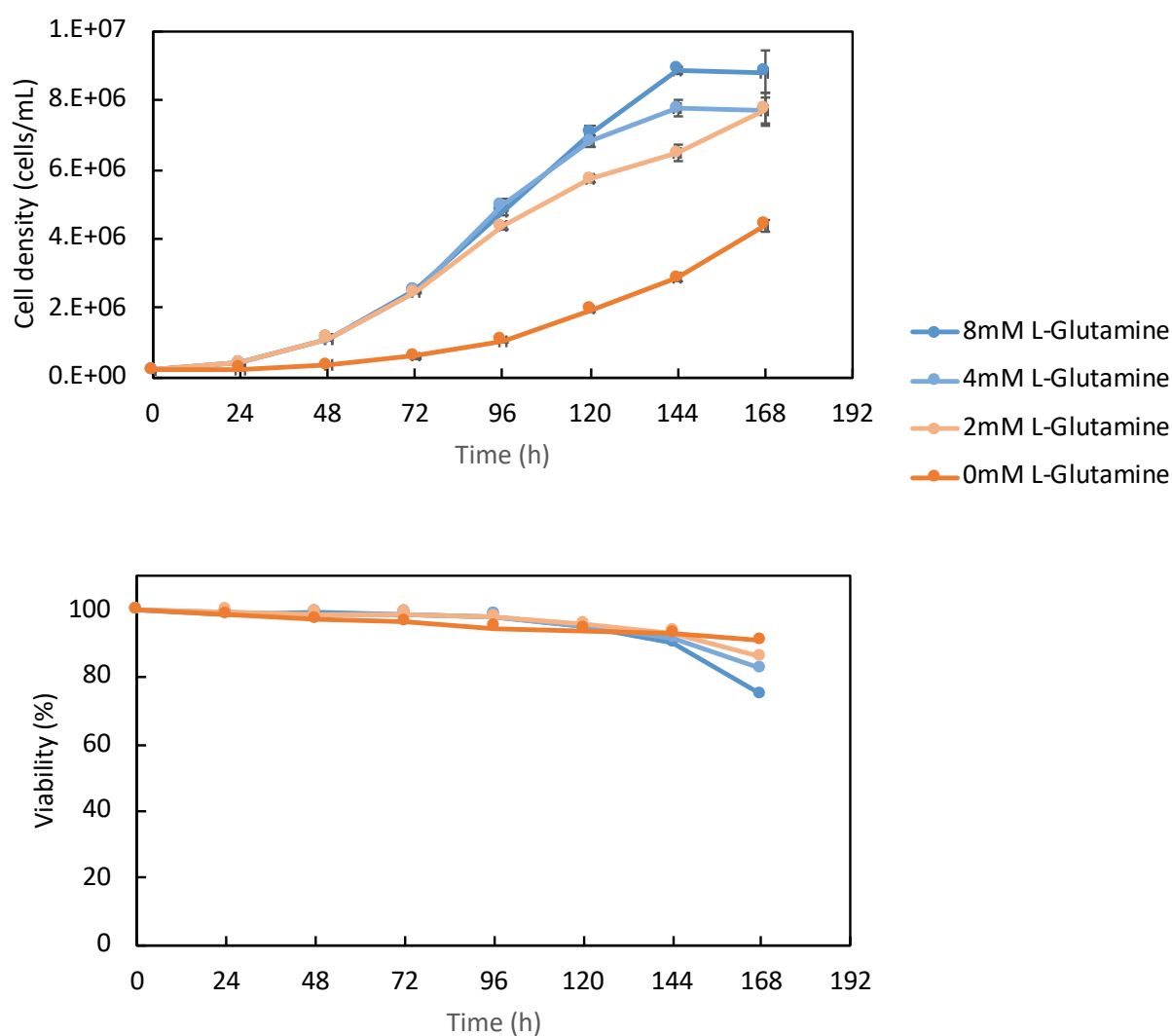
Supplementary Figure 5 Effect of reducing L-glutamine concentration in the media on growth and viability of the parental CHO-K1. Cells were seeded BalanCD™ media containing different amounts of L-glutamine ranging from 0 mM to 8 mM in Tubespin® 50 bioreactors (TPP) with a 5 mL working volume. Cell density and viability were monitored using the GUAVA EasyCyte ™ Viacount reagent. Each data point represents the average of 3 replicate cultures and error bars represent standard deviation.

Time (h)	GN (h^{-1})	NN (h^{-1})
24	0.027	0.019
48	0.026	0.034
72	0.040	0.035
96	0.028	0.028
120	0.022	0.020
144	0.007	0.017
168	0.000	0.007
192	-0.024	-0.031

Supplementary Figure 6 Specific growth rates calculated for the data shown in Figure 1B. Parental CHO-K1 in BalanCD media supplemented with 8 mM L-glutamine (GN) were compared to glutamine-free adapted CHO-K1 (NN) grown in BalanCD media not supplemented with L-glutamine. Specific growth rates are calculated as described in Methods
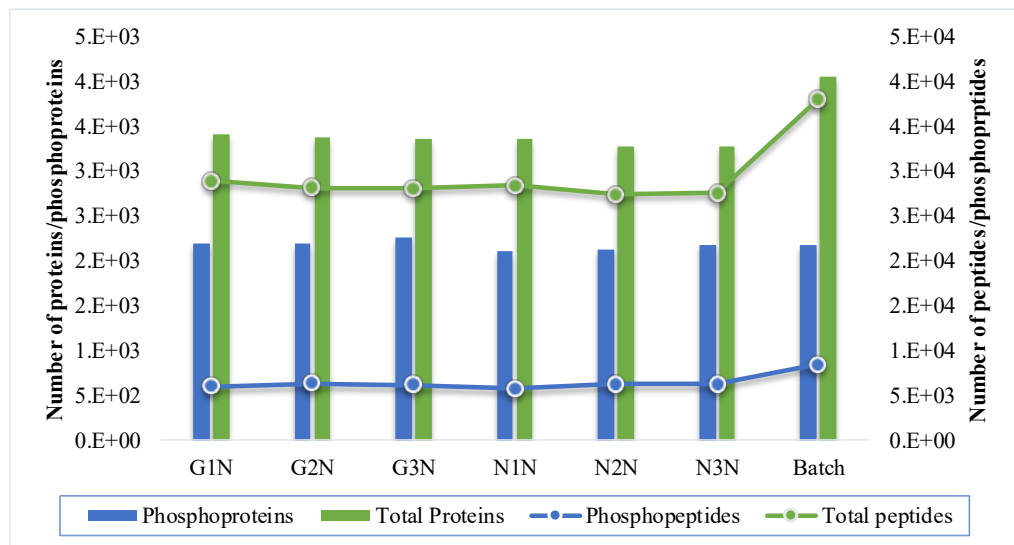
Supplementary Figure 7 Summary of total proteomic identifications. The Bar chart refers to the primary y-axis (on the left) that represent the total number of phosphoproteins (Blue) and total proteins with the number of unique peptides ≥2 (Green) identified from three biological replicates of condition GN and NN. The line graph refers to the secondary y-axis (on the right) that displayed the number of peptides identified from the non-phospho-enriched fractions

(Blue) and the number of phosphopeptides (PhosphoRS Best Site Probability ≥ 75%) identified from total protein fractions (Green)

A) Overlap In protein identification
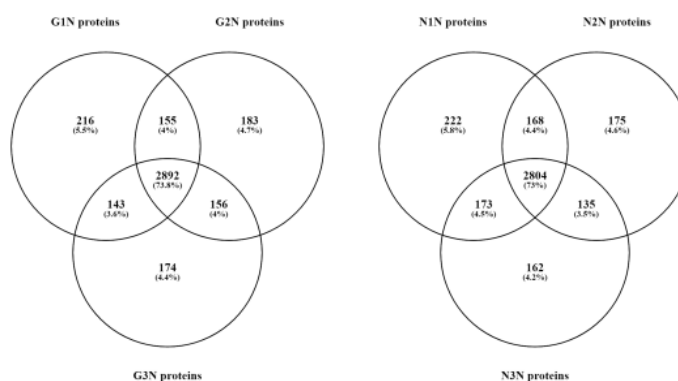
B) Overlap In phosphoprotein identification
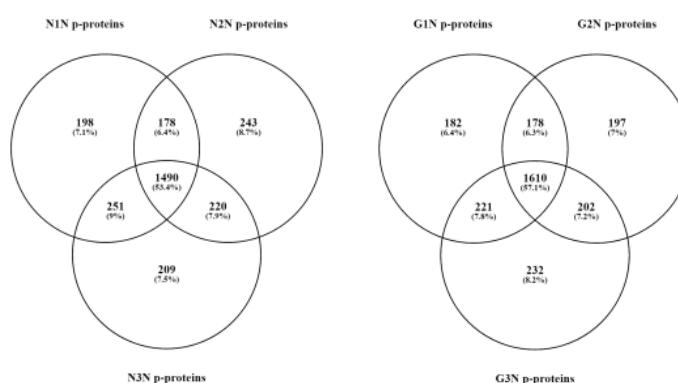
Supplementary Figure 8 Venn diagrams of overlapping identifications. Shared and unique protein (A) and phosphoprotein (B) identifications among biological replicates of GN and and NN sample groups.

Supplementary tables can be downloaded from

https://drive.google.com/open?id=1XskMA0UiS2oHTOZkaGyAob65nbUihR_G

5. Chapter 5 - Future Work and Conclusions

Advancements in mass spectrometric phosphoproteomics have revolutionized cell signalling research, especially in understanding cancer (Singh et al., 2017). The application of phosphoproteomics towards understanding CHO cells biology is relatively new. This thesis work involved the incorporation of global phosphoproteomics research into the understanding of CHO cell biology in bioprocess-relevant phenotypes and actively participated in four out of the five large scale CHO phosphoproteomics studies published so far, three of which makes the core body of this thesis. These studies explore three critical attributes of CHO cell biology, i.e., growth in Chapter 2, productivity in Chapter 3 and response to nutrient limitation in Chapter 4, and reveal crucial information about proteins and phosphoproteins involved in regulating these complex biological processes. Each of the studies has the potential for further investigation, which is discussed below. But overall, these studies also share a similar technical background and therefore, a similar share of technical limitations. The future work, therefore, will be focused on expanding the scope and reach of these studies by;

1- **Increasing the number of confident phosphosites identified using a combination on phosphopeptide enrichment methods**: The studies discussed in this thesis use only one phosphopeptide enrichment technique and studies have shown that using multiple phosphopeptide enrichments techniques increase the phosphoproteome coverage (Henry, Power, et al., 2017b). For example, sequential elution from IMAC (SIMAC) method first elutes multi-phosphorylated phosphopeptides using IMAC chromatography (used in chapter 2) followed by mono-phosphorylated phosphopeptide enrichment by TiO_2 (used in Chapter 3 and 4) thus removing the bias in

phosphopeptide enrichment and ultimately increasing the number of phosphopeptides identified (Thingholm & Larsen, 2016b). Pre fractionation of whole-cell peptide mixture using separation techniques such as strong cation exchange (SCX), strong anion exchange (SAX), and C18 material for basic reversed-phase (BRP) fractionation before phosphopeptide enrichment has also shown increased phosphopeptide coverage (Dehghani, Gödderz, & Winter, 2018). Future work on CHO phosphoproteomic would use the combination of enrichment techniques to improve the phosphoproteome coverage and depth.

2- **Parallel enrichment for peptides phosphorylated at tyrosine residue:** The Underrepresentation of phosphotyrosine (p-Tyr) residues in the phospho-enriched peptide pool is a severe limitation with conventional chromatographic techniques such as IMAC and TiO_2 used in this thesis. And given the conserved evolutionary nature of tyrosine phosphorylation (Tong, Wang, & Su, 2017) and its interaction with Src homology (SH) 2 domains and pTyr-binding (PTB) domains (L. Li et al., 2012), the future work on CHO cell phosphoproteomic will focus on selective enrichment of tyrosine-phosphorylated peptides by capturing p-Tyr restudies by immobilized anti-P-Tyr antibodies (Bllaci et al., 2017).

3- **Absolute quantification of phosphorylation:** The studies presented in this thesis utilise relative qualification techniques to decipher the differential expression of phosphorylated peptide/phosphoprotein between sample types. Although relative quantification archives substantial phosphoproteome coverage, it is biased towards the identification of high abundant phosphopeptides (Michalski, Cox, & Mann, 2011) and the stoichiometry of phosphorylated peptide to the total peptide remains unknown. Current state-

of-the-art mass spectrometers such Orbitrap Fusion™ Tribrid™ Mass Spectrometers are capable of selecting pre-defined peptides for absolute quantification using Parallel reaction monitoring (PRM) or Selective reaction monitoring (SRM) based techniques by isolating target precursor ions and acquiring high-resolution spectra with high reproducibility (Picotti et al., 2010). A recent study used PRM was used to select specific phosphopeptides for absolute quantification along with its non-phosphorylated isoform and precisely determined the phosphorylation percentage of the specific phosphosite (Dekker et al., 2018), thus providing a more realistic measure of kinase/phosphatase activity. The studies presented in this thesis offers several critical phosphosites as exciting candidates for absolute quantification.

4- **Functional annotation of phosphosites.** A typical phosphoproteomic experiment reveals thousands of phosphorylation sites with unknown practical relevance. A decade of phosphoproteomic research in cancer has reached a stage of prioritising functionally relevant phosphosites with the use of machine learning approaches for interrogation using genome editing techniques such as CRISPR/Cas9 and targeted genome editing (Ochoa et al., 2019). A similar approach with CHO cell phosphoproteomics could reveal novel phosphosites for relevance in cell line engineering.

5- **Targeting phosphosites to achieve the desired phenotype.** Functional annotation of the phosphorylation status of a phosphosite will enable the site-specific targeting by kinase inhibitors or genome-editing techniques such as CRISPR-Cas9 and site-directed mutagenesis to achieve gain or loss-of-function mutations and resultant phenotypic changes. Chapter 2, for example, identified differential expression of 94 kinases in the context of CHO cell

growth, including eight cyclin-dependent kinases (Cdks). CdKs have been previously targeted in CHO cells by small-molecule inhibitor compound to achieve growth arrest and an increase in specific productivity (Du et al., 2015). This thesis work has provided a substantial increase in the identification of bioprocess relevant kinases and phosphosites in CHO cells, thus increasing the scope for the application of pathway manipulation towards achieving desirable CHO phenotypes.

The future strategies described above are common to all three studies discussed in this thesis whereas each study also has the potential for further scientific investigation

Chapter 2- "*The Expression Pattern of the Phosphoproteome Is Significantly Changed During the Growth Phases of Recombinant CHO Cell Culture*". This study explores the temporal changes in the proteomic and phosphoproteomic landscape of CHO cell during growth in suspension batch culture. Through the identification of differential proteins and phosphoproteins, this study discusses critical proteins and pathways such as the mTOR pathway as potential targets for developing growth control strategies using cell line engineering or kinase inhibition via small molecule inhibitors. Although this study provides information about the dynamics of CHO proteome and phosphoproteome in response to growth, this study suffers from several limitations:

1- 4-day cell culture in 35ml shake flasks is far from the scale at which industrial bioprocess is performed and therefore is not an ideal representation of cellular dynamics at production levels.

2- In this study, only the proteomic and phosphoproteomic changes are accounted for CHO cell growth. However, cellular growth is a complex biological process, and similar studies concerning other regulatory mechanisms have

shown differences in mRNA and miRNA expression (Fomina-Yadlin et al., 2015), epigenetic modification of cellular DNA (Hernandez et al., 2019), differential gene expression (Reinhart, Damjanovic, Castan, Ernst, & Kunert, 2018b) associated with CHO cell growth. Therefore, to comprehensively understand CHO cell growth, other regulatory mechanisms should be studied in conjunction.

A solution to these limitations is a large-scale experimental setup using industrial CHO cell line cultivated in the near-bioproduction setting to mimic the actual industrial recombinant protein production process. The samples collected from this will be used to perform multiple large-scale omics techniques with the ultimate aim to integrate all the omics data and provide a comprehensive database for CHO cell growth dynamics. Large-scale studies like these, however, are rare in academic settings due to time, infrastructure and resource constraints.

Fortunately, this PhD project is part of the Marie-curie, eCHO Systems international training network (ITN), which is a collaborative effort of four academic institutes across Europe, training 15 PhD students towards a collective goal of improving CHO cell platform for recombinant protein production. Through industrial collaboration with the UCB pharmaceuticals in Belgium, two 10L scale bioreactors were cultured with one of their production DG44 CHO cell line. Samples collected from these bioreactors were sent to eCHO collaborative partners based on their area of expertise for omics analysis. Table 5.1 provides a list of proposed and ongoing investigation for this project.

Table 5.1 eCHO-ITN, proposed multi-omics analysis.

Techniques	Proposed analysis	Institute
Proteomics	Quantitative label-free LC-MS/MS differential analysis	DCU
Phospho-proteomics	Phosphopeptide enrichment followed by quantitative label-free LC-MS/MS differential analysis	DCU
tRNA analysis	Identification of tRNAs, miRNAs, snRNAs	University of Kent
Total RNAseq	Identification of transcripts, isoforms and lncRNAs	NIBRT
Glycan	The N-glycan pattern of the secreted proteins and N-Glycan profile of purified secreted antibody	Denmark Technical University
Genomics	Sequencing of the cell's genome and methylome analysis	BOKU- Vienna

Chapter 3- "*Increased mAb production in amplified CHO cell lines is associated with increased interaction of CREB1 with transgene promoter*". While Chapter 3 aims to explicitly enquire the changes in expression, modification, and binding of transcription factors as an explanation for differences in productivity state between the two cell lines. Therefore, to increase the identification of transcription factors, nuclear proteomic and phosphoproteomic analysis was performed for the samples collected at the exponential phase, i.e., the period of most significant productivity difference between the two cell lines. Future work will explore overall changes in cellular regulation in response to increased productivity. A more extensive comparison of transcriptomic, proteomic and phosphoproteomic differences between the two cell lines in both exponential and the stationary phase is the subject of a manuscript in preparation.

Chapter 4- "*Quantitative label-free LC-MS/MS comparative proteomic and phosphoproteomic analysis of CHO-K1 cells adapted to growth in glutamine-free media.*" By studying the changes in proteome and phosphoproteome of CHO-K1 cells adapted to growth in glutamine free media, this study opens new possibilities of characterising CHO cell response to nutrient limitation. A recent study showed a varied response in the CHO cell culture performance subjected to restriction to different amino acids (Ghaffari et al., 2019), it would be interesting to study these

changes through phosphoproteomics to identify specific pathways affected by particular amino acid limitation.

5.1 Concluding remarks

Chinese hamster ovary cells are an industrial workhorse to produce recombinant protein products, and the last decade has seen remarkable improvements in the ability of these cells to produce high-quality therapeutic protein product in greater quantities. However, the CHO cell system is far from ideal; several limitations such as uncontrolled growth, low productivity and toxic by-product formation are keeping these cells away from their true potential. Advancements in media design, bioprocess optimisation and other iterative processes have already pushed the cells to the limits, and now, it is the biology of CHO cells that needs to be channelled towards producing more efficacious protein products in higher quantities. By taking one step at a time, cell line engineering approaches have explored the possibilities of making changes in CHO biological processes toward achieving desired bioprocess phenotypes. Recent advancements in genome-scale modelling and artificial intelligence-based predictions have the potential to accelerate this growth, fundamental to which is an increased understanding of how complex bioprocesses such as growth, protein synthesis, metabolism are regulated in CHO cells. The majority of CHO cell research towards understanding these biological processes is performed on genomic and transcriptomic control of CHO cells. Proteins, however, are the final effector and regulatory molecules and the cellular state is controlled by maintaining the correct stoichiometry of proteins, and the changes in protein expression are often linked to changing cellular states. Therefore, studying protein expression and the change in protein expression under bioprocess conditions is likely to provide a more realistic account of CHO cell biology under specific conditions. The activity of proteins is also regulated by post-

translational modification and therefore studying the dynamics of protein PTMs is equally essential.

Post-translational modification of proteins by reversible phosphorylation of serine, threonine and tyrosine residues is the most common type of protein activity control, and due to the ubiquitous nature phosphorylation, it affects almost all cellular processes including the processes that are highly relevant in bioprocessing. Given the importance of phosphorylation in biological regulation through cell signalling, it is surprising that it has not been extensively studied in CHO cells. This thesis identifies this knowledge gap and performs phosphoproteomic research towards understanding three important bioprocess attributes of CHO cells, i.e., growth, productivity and metabolism. Chapter 2 for the first time shows the dynamic nature of the CHO cell phosphoproteome by identifying differential phosphorylation of 1415 phosphoproteins as cell culture progresses from one phase of growth to the other. These phosphoproteins include 94 kinases that can be targeted by genome editing technique or small molecule inhibitors to modulate CHO cell growth in a bioreactor. Chapter 3 identifies the activation of transcription factor CREB1 as a potential explanation for increased protein productivity in a high producer cell line. This study further stresses the importance of studying proteomic and phosphoproteomic, along with transcriptomic changes in CHO cells. Although the differential mRNA analysis between high producer A1 cell line and the low producer A0 cell line predicted increased activity of CREB1 transcription factor in A1 cells, the proteomic analysis of the nuclear proteome of these cells showed decreased expression of CREB1 protein in A1 cells. However, by studying the nuclear phosphoproteome, we were also to identify phosphorylation pattern on CREB1 protein in A1 cell that indicates CREB1 activation which was later confirmed by ChIP analysis which showed 6-fold increased

binding of the CREB1 to the CMV promotor. Finally, Chapter 4 provides a proteomic and phosphoproteomic response to long term glutamine starvation of CHO K1 cells. Glutamine supplementation has been an active area of research in CHO, and although studies have previously adapted CHO cells to growth in glutamine free media, we provide the first account of global proteomic and phosphoproteomic differences resulting from this adaptation. The functional analysis of this data shows differential expression of proteins and phosphoproteins from mRNA processing via the spliceosome, and cytoskeleton rearrangement.

6. References

Abelin, J. G., Patel, J., Lu, X., Feeney, C. M., Fagbami, L., Creech, A. L., … Jaffe, J. D. (2016). Reduced-representation Phosphosignatures Measured by Quantitative Targeted MS Capture Cellular States and Enable Large-scale Comparison of Drug-induced Phenotypes. *Molecular & Cellular Proteomics*, *15*(5), 1622–1641. https://doi.org/10.1074/mcp.M116.058354

Aguilera, M. O., Berón, W., & Colombo, M. I. (2012). The actin cytoskeleton participates in the early events of autophagosome formation upon starvation induced autophagy. *Autophagy*, *8*(11), 1590–1603. https://doi.org/10.4161/auto.21459

Akopov, S. B., Ruda, V. M., Batrak, V. V, Vetchinova, A. S., Chernov, I. P., Nikolaev, L. G., … Sverdlov, E. D. (2006). Identification, genome mapping, and CTCF binding of potential insulators within the FXYD5-COX7A1 locus of human chromosome 19q13.12. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, *17*(10), 1042–1049. https://doi.org/10.1007/s00335-006-0037-3

Alers, S., Loffler, A. S., Wesselborg, S., & Stork, B. (2012). Role of AMPK-mTOR-Ulk1/2 in the Regulation of Autophagy: Cross Talk, Shortcuts, and Feedbacks. *Molecular and Cellular Biology*, *32*(1), 2–11. https://doi.org/10.1128/MCB.06159-11

Altamirano, C., Paredes, C., Cairó, J. J., & Gòdia, F. (2000). Improvement of CHO cell culture medium formulation: Simultaneous substitution of glucose and glutamine. *Biotechnology Progress*, *16*(1), 69–75. https://doi.org/10.1021/bp990124j

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, G. L., Dean, R. A., Hawkridge, A. M., & Muddiman, D. C. (2011). Improving proteome coverage on a LTQ-orbitrap using design of experiments. *Journal of the American Society for Mass Spectrometry*, *22*(4), 773–783. https://doi.org/10.1007/s13361-011-0075-2

Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W., & Smith, R. D. (2012). Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society Reviews*, *41*(10), 3912. https://doi.org/10.1039/c2cs15331a

Aoki, K., Yamada, M., Kunida, K., Yasuda, S., & Matsuda, M. (2011). Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proceedings of the National Academy of Sciences*, *108*(31), 12675–12680. https://doi.org/10.1073/pnas.1104030108

Ardito, F., Giuliani, M., Perrone, D., Troiano, G., & Lo Muzio, L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *International Journal of Molecular Medicine*. https://doi.org/10.3892/ijmm.2017.3036

Aryal, U. K., & Ross, A. R. S. (2010). Enrichment and analysis of phosphopeptides under different experimental conditions using titanium dioxide affinity chromatography and mass spectrometry. *Rapid Communications in Mass Spectrometry*, *24*(2), 219–231. https://doi.org/10.1002/rcm.4377

Baek, E., Kim, C. L., Kim, M. G., Lee, J. S., & Lee, G. M. (2016). Chemical inhibition of autophagy: Examining its potential to increase the specific productivity of recombinant CHO cell lines. *Biotechnology and Bioengineering*, *113*(9), 1953–1961. https://doi.org/10.1002/bit.25962

Baik, J. Y., Lee, M. S., An, S. R., Yoon, S. K., Joo, E. J., Kim, Y. H., … Lee, G. M. (2006). Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin. *Biotechnology and Bioengineering*. https://doi.org/10.1002/bit.20717

Bandaranayake, A. D., & Almo, S. C. (2014). Recent advances in mammalian protein production. *FEBS Letters*. https://doi.org/10.1016/j.febslet.2013.11.035

Bauman, A. L., & Scott, J. D. (2002). Kinase- and phosphatase-anchoring proteins: harnessing the dynamic duo. *Nature Cell Biology*, *4*(8), E203-6. https://doi.org/10.1038/ncb0802-e203

Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villén, J., Li, J., … Gygi, S. P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proceedings of the National*

Academy of Sciences of the United States of America, *101*(33), 12130–12135. https://doi.org/10.1073/pnas.0404720101

Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J., & Gygi, S. P. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol*, *24*(10), 1285–1292. https://doi.org/10.1038/nbt1240

Becker, J., Hackl, M., Rupp, O., Jakobi, T., Schneider, J., Szczepanowski, R., … Brinkrolf, K. (2011). Unraveling the Chinese hamster ovary cell line transcriptome by next-generation sequencing. *Journal of Biotechnology*, *156*(3), 227–235. https://doi.org/10.1016/j.jbiotec.2011.09.014

Becker, M. M. M., Lapouge, K., Segnitz, B., Wild, K., & Sinning, I. (2017). Structures of human SRP72 complexes provide insights into SRP RNA remodeling and ribosome interaction. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkw1124

Bedoya-López, A., Estrada, K., Sanchez-Flores, A., Ramírez, O. T., Altamirano, C., Segovia, L., … Valdez-Cruz, N. A. (2016). Effect of Temperature Downshift on the Transcriptomic Responses of Chinese Hamster Ovary Cells Using Recombinant Human Tissue Plasminogen Activator Production Culture. *PLOS ONE*, *11*(3), e0151529. https://doi.org/10.1371/journal.pone.0151529

Beltran, L., & Cutillas, P. R. (2012). Advances in phosphopeptide enrichment techniques for phosphoproteomics. *Amino Acids*. https://doi.org/10.1007/s00726-012-1288-9

Bhullar, K. S., Lagarón, N. O., McGowan, E. M., Parmar, I., Jha, A., Hubbard, B. P., & Rupasinghe, H. P. V. (2018). Kinase-targeted cancer therapies: progress, challenges and future directions. *Molecular Cancer*, *17*(1), 48. https://doi.org/10.1186/s12943-018-0804-2

Black, D. L. (2003). Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*, *72*(1), 291–336. https://doi.org/10.1146/annurev.biochem.72.121801.161720

Bllaci, L., Torsetnes, S. B., Wierzbicka, C., Shinde, S., Sellergren, B., Rogowska-Wrzesinska, A., & Jensen, O. N. (2017). Phosphotyrosine Biased Enrichment of Tryptic Peptides from Cancer Cells by Combining pY-MIP and TiO2 Affinity Resins. *Analytical Chemistry*, *89*(21), 11332–11340. https://doi.org/10.1021/acs.analchem.7b02091

Bollati-Fogolín, M., Forno, G., Nimtz, M., Conradt, H. S., Etcheverrigaray, M., & Kratje, R. (2005). Temperature reduction in cultures of hGM-CSF-expressing CHO cells: Effect on productivity and product quality. *Biotechnology Progress*, *21*(1), 17–21. https://doi.org/10.1021/bp049825t

Bondarenko, P. V., Chelius, D., & Shaler, T. A. (2002). Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography - Tandem mass spectrometry. *Analytical Chemistry*, *74*(18), 4741–4749. https://doi.org/10.1021/ac0256991

Bort, J. A. H., Hackl, M., Höflmayer, H., Jadhav, V., Harreither, E., Kumar, N., … Borth, N. (2012). Dynamic mRNA and miRNA profiling of CHO-K1 suspension cell cultures. *Biotechnology Journal*, *7*(4), 500–515. https://doi.org/10.1002/biot.201100143

Bort, J. A. H., Stern, B., & Borth, N. (2010a). CHO-K1 host cells adapted to growth in glutamine-free medium by FACS-assisted evolution. *Biotechnology Journal*, *5*(10), 1090–1097. https://doi.org/10.1002/biot.201000095

Bort, J. A. H., Stern, B., & Borth, N. (2010b). CHO-K1 host cells adapted to growth in glutamine-free medium by FACS-assisted evolution. *Biotechnology Journal*, *5*(10), 1090–1097. https://doi.org/10.1002/biot.201000095

Braiman, A., & Isakov, N. (2015). The role of Crk adaptor proteins in T-cell adhesion and migration. *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2015.00509

Brinkrolf, K., Rupp, O., Laux, H., Kollin, F., Ernst, W., Linke, B., … Borth, N. (2013). Chinese hamster genome sequenced from sorted chromosomes. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2645

Brown, A. J., Sweeney, B., Mainwaring, D. O., & James, D. C. (2014). Synthetic promoters for CHO cell engineering. *Biotechnology and Bioengineering*, *111*(8), 1638–1647. https://doi.org/10.1002/bit.25227

Brown, A. J., Sweeney, B., Mainwaring, D. O., & James, D. C. (2015). NF-κB, CRE and YY1 elements

are key functional regulators of CMV promoter-driven transient gene expression in CHO cells. *Biotechnology Journal*, *10*(7), 1019–1028. https://doi.org/10.1002/biot.201400744

Bylund, D., Danielsson, R., Malmquist, G., & Markides, K. E. (2002). Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modelling of liquid chromatography-mass spectrometry data. *Journal of Chromatography A*, *961*(2), 237–244. https://doi.org/10.1016/S0021-9673(02)00588-5

Caenepeel, S., Charydczak, G., Sudarsanam, S., Hunter, T., & Manning, G. (2004). The mouse kinome: Discovery and comparative genomics of all mouse protein kinases. *Proceedings of the National Academy of Sciences*, *101*(32), 11707–11712. https://doi.org/10.1073/pnas.0306880101

Cantin, G. T., Shock, T. R., Sung, K. P., Madhani, H. D., & Yates, J. R. (2007). Optimizing TiO2-based phosphopeptide enrichment for automated multidimensional liquid chromatography coupled to tandem mass spectrometry. *Analytical Chemistry*, *79*(12), 4666–4673. https://doi.org/10.1021/ac0618730

Chen, J., Bardes, E. E., Aronow, B. J., & Jegga, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkp427

Chen, L., Chen, D. T., Chen, T., Kurtyka, C., Rawal, B., Fulp, W. J., … Cress, W. D. (2012). Tripartite motif containing 28 (Trim28) can regulate cell proliferation by bridging HDAC1/E2F interactions. *Journal of Biological Chemistry*, *287*(48), 40106–40118. https://doi.org/10.1074/jbc.M112.380865

Cheung, R. C. F., Wong, J. H., & Ng, T. B. (2012). Immobilized metal ion affinity chromatography: A review on its applications. *Applied Microbiology and Biotechnology*. https://doi.org/10.1007/s00253-012-4507-0

Chiang, A. W. T., Li, S., Kellman, B. P., Chattopadhyay, G., Zhang, Y., Kuo, C. C., … Lewis, N. E. (2019). Combating viral contaminants in CHO cells by engineering innate immunity. *Scientific Reports*, *9*(1), 8827. https://doi.org/10.1038/s41598-019-45126-x

Choudhary, C., & Mann, M. (2010). Decoding signalling networks by mass spectrometry-based proteomics. *Nature Reviews Molecular Cell Biology*, *11*(6), 427–439. https://doi.org/10.1038/nrm2900

Clark, M. D., Kumar, G. S., Marcum, R., Luo, Q., Zhang, Y., & Radhakrishnan, I. (2015). Molecular Basis for the Mechanism of Constitutive CBP/p300 Coactivator Recruitment by CRTC1-MAML2 and Its Implications in cAMP Signaling. *Biochemistry*. https://doi.org/10.1021/acs.biochem.5b00332

Cluntun, A. A., Lukey, M. J., Cerione, R. A., & Locasale, J. W. (2017). Glutamine Metabolism in Cancer: Understanding the Heterogeneity. *Trends in Cancer*, *3*(3), 169–180. https://doi.org/10.1016/j.trecan.2017.01.005

Cohen, P. (2000). The regulation of protein function by multisite phosphorylation--a 25 year update. *Trends in Biochemical Sciences*, *25*(12), 596–601. https://doi.org/10.1016/S0968-0004(00)01712-6

Cohen, Philip. (2000). The regulation of protein function by multisite phosphorylation - A 25 year update. *Trends in Biochemical Sciences*, *25*(12), 596–601. https://doi.org/10.1016/S0968-0004(00)01712-6

Dadehbeigi, N., & Dickson, A. J. (2015). Chemical manipulation of the mTORC1 pathway in industrially relevant CHOK1 cells enhances production of therapeutic proteins. *Biotechnology Journal*, *10*(7), 1041–1050. https://doi.org/10.1002/biot.201500075

Dahodwala, H., Kaushik, P., Tejwani, V., Kuo, C.-C., Menard, P., Henry, M., … Sharfstein, S. T. (2019a). Increased mAb production in amplified CHO cell lines is associated with increased interaction of CREB1 with transgene promoter. *Current Research in Biotechnology*, *1*, 49–57. https://doi.org/10.1016/j.crbiot.2019.09.001

Dahodwala, H., Kaushik, P., Tejwani, V., Kuo, C.-C., Menard, P., Henry, M., … Sharfstein, S. T. (2019b). Increased mAb production in amplified CHO cell lines is associated with increased interaction of CREB1 with transgene promoter. *Current Research in Biotechnology*, *1*, 49–57.

https://doi.org/10.1016/j.crbiot.2019.09.001

Dahodwala, H., Nowey, M., Mitina, T., & Sharfstein, S. T. (2012). Effects of clonal variation on growth, metabolism, and productivity in response to trophic factor stimulation: A study of Chinese hamster ovary cells producing a recombinant monoclonal antibody. *Cytotechnology*, *64*(1), 27–41. https://doi.org/10.1007/s10616-011-9388-z

Dahodwala, H., & Sharfstein, S. T. (2014). Role of epigenetics in expression of recombinant proteins from mammalian cells. *Pharmaceutical Bioprocessing*, *2*(5), 403–419. https://doi.org/10.4155/pbp.14.47

Dahodwala, H., & Sharfstein, S. T. (2017a). Biosimilars: Imitation Games. *ACS Medicinal Chemistry Letters*, *8*(7), 690–693. https://doi.org/10.1021/acsmedchemlett.7b00199

Dahodwala, H., & Sharfstein, S. T. (2017b). The 'Omics Revolution in CHO Biology: Roadmap to Improved CHO Productivity BT - Heterologous Protein Production in CHO Cells: Methods and Protocols. In P. Meleady (Ed.) (pp. 153–168). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-6972-2_10

Datta, P., Linhardt, R. J., & Sharfstein, S. T. (2013). An 'omics approach towards CHO cell engineering. *Biotechnology and Bioengineering*. https://doi.org/10.1002/bit.24841

Decker, T., & Kovarik, P. (2000). Serine phosphorylation of STATs. *Oncogene*, *19*(21), 2628–2637. https://doi.org/10.1038/sj.onc.1203481

DeGnore, J., & Qin, J. (1998). Fragmentation of phosphopeptides in an ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry*, *9*(11), 1175–1188. https://doi.org/10.1016/S1044-0305(98)00088-9

Dehghani, A., Gödderz, M., & Winter, D. (2018). Tip-Based Fractionation of Batch-Enriched Phosphopeptides Facilitates Easy and Robust Phosphoproteome Analysis. *Journal of Proteome Research*, *17*(1), 46–54. https://doi.org/10.1021/acs.jproteome.7b00256

Dekker, L. J. M., Zeneyedpour, L., Snoeijers, S., Joore, J., Leenstra, S., & Luider, T. M. (2018). Determination of Site-Specific Phosphorylation Ratios in Proteins with Targeted Mass Spectrometry. *Journal of Proteome Research*, *17*(4), 1654–1663. https://doi.org/10.1021/acs.jproteome.7b00911

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dorner, A. J., Wasley, L. C., & Kaufman, R. J. (1989). Increased synthesis of secreted proteins induces expression of glucose-regulated proteins in butyrate-treated Chinese hamster ovary cells. *Journal of Biological Chemistry*.

Dreesen, I. A. J., & Fussenegger, M. (2011). Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells. *Biotechnology and Bioengineering*, *108*(4), 853–866. https://doi.org/10.1002/bit.22990

Druz, A., Son, Y. J., Betenbaugh, M., & Shiloach, J. (2013). Stable inhibition of mmu-miR-466h-5p improves apoptosis resistance and protein production in CHO cells. *Metabolic Engineering*, *16*(1), 87–94. https://doi.org/10.1016/j.ymben.2012.12.004

Du, Z., Treiber, D., Mccarter, J. D., Fomina-Yadlin, D., Saleem, R. A., Mccoy, R. E., … Reddy, P. (2015). Use of a small molecule cell cycle inhibitor to control cell growth and improve specific productivity and product quality of recombinant proteins in CHO cell cultures. *Biotechnology and Bioengineering*, *112*(1), 141–155. https://doi.org/10.1002/bit.25332

Dumont, J., Euwart, D., Mei, B., Estes, S., & Kshirsagar, R. (2016). Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical Reviews in Biotechnology*. https://doi.org/10.3109/07388551.2015.1084266

Duong-Ly, K. C., & Peterson, J. R. (2013). The human kinome and kinase inhibition. *Current Protocols in Pharmacology*, (SUPPL.60). https://doi.org/10.1002/0471141755.ph0209s60

El-Faramawy, A., Siu, K. W. M., & Thomson, B. A. (2005). Efficiency of nano-electrospray ionization. *Journal of the American Society for Mass Spectrometry*, *16*(10), 1702–1707.

https://doi.org/10.1016/j.jasms.2005.06.011

Eng, J. K., Mccormack, A. L., & Yates, J. R. (1994). An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *American Society for Mass Spectrometry*, *5*, 976–989. https://doi.org/10.1016/1044-0305(94)80016-2

Eng, J. K., McCormack, A. L., & Yates, J. R. (1994). 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*, *5*(11), 976–989. https://doi.org/10.1016/1044-0305(94)80016-2

Erickson, B. K., Jedrychowski, M. P., McAlister, G. C., Everley, R. A., Kunz, R., & Gygi, S. P. (2015). Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer. *Analytical Chemistry*, *87*(2), 1241–1249. https://doi.org/10.1021/ac503934f

Evnin, L. B., Vasquez, J. R., & Craik, C. S. (1990). Substrate specificity of trypsin investigated by using a genetic selection. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(17), 6659–6663. https://doi.org/10.1073/pnas.87.17.6659

Fallot, S., Ben Naya, R., Hieblot, C., Mondon, P., Lacazette, E., Bouayadi, K., … Prats, H. (2009). Alternative-splicing-based bicistronic vectors for ratio-controlled protein expression and application to recombinant antibody production. *Nucleic Acids Research*, *37*(20), e134–e134. https://doi.org/10.1093/nar/gkp716

Feichtinger, J., Hernández, I., Fischer, C., Hanscho, M., Auer, N., Hackl, M., … Borth, N. (2016). Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time. *Biotechnology and Bioengineering*, *113*(10), 2241-2253. https://doi.org/10.1002/bit.25990

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*, *246*(4926), 64–71. https://doi.org/10.1126/science.2675315

Ferries, S., Perkins, S., Brownridge, P. J., Campbell, A., Eyers, P. A., Jones, A. R., & Eyers, C. E. (2017). Evaluation of Parameters for Confident Phosphorylation Site Localization Using an Orbitrap Fusion Tribrid Mass Spectrometer. *Journal of Proteome Research*, *16*(9), 3448–3459. https://doi.org/10.1021/acs.jproteome.7b00337

Fílla, J., & Honys, D. (2012). Enrichment techniques employed in phosphoproteomics. *Amino Acids*. https://doi.org/10.1007/s00726-011-1111-z

Fischer, S., Handrick, R., & Otte, K. (2015). The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnology Advances*. https://doi.org/10.1016/j.biotechadv.2015.10.015

Fomina-Yadlin, D., Mujacic, M., Maggiora, K., Quesnell, G., Saleem, R., & McGrew, J. T. (2015). Transcriptome analysis of a CHO cell line expressing a recombinant therapeutic protein treated with inducers of protein expression. *Journal of Biotechnology*, *212*, 106–115. https://doi.org/10.1016/j.jbiotec.2015.08.025

Forgacs, G., Yook, S. H., Janmey, P. A., Jeong, H., & Burd, C. G. (2004). Role of the cytoskeleton in signaling networks. *Journal of Cell Science*, *117*(13), 2769–2775. https://doi.org/10.1242/jcs.01122

Foster, K. G., Acosta-Jaquez, H. A., Romeo, Y., Ekim, B., Soliman, G. A., Carriere, A., … Fingar, D. C. (2010). Regulation of mTOR complex 1 (mTORC1) by raptor Ser863 and multisite phosphorylation. *Journal of Biological Chemistry*, *285*(1), 80–94. https://doi.org/10.1074/jbc.M109.029637

Frame, M. C. (2004). Newest findings on the oldest oncogene; how activated src does it. *Journal of Cell Science*, *117*(7), 989–998. https://doi.org/10.1242/jcs.01111

Frese, C. K., Zhou, H., Taus, T., Altelaar, A. F. M., Mechtler, K., Heck, A. J. R., & Mohammed, S. (2013). Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *Journal of Proteome Research*, *12*(3), 1520–1525. https://doi.org/10.1021/pr301130k

Ganley, I. G., Lam, D. H., Wang, J., Ding, X., Chen, S., & Jiang, X. (2009). ULK1·ATG13·FIP200

complex mediates mTOR signaling and is essential for autophagy. *Journal of Biological Chemistry*, *284*(18), 12297–12305. https://doi.org/10.1074/jbc.M900573200

Gara, R. K., Kumari, S., Ganju, A., Yallapu, M. M., Jaggi, M., & Chauhan, S. C. (2015). Slit/Robo pathway: A promising therapeutic target for cancer. *Drug Discovery Today*, *20*(1), 156–164. https://doi.org/10.1016/j.drudis.2014.09.008

Ge, Q.-L., Liu, S.-H., Ai, Z.-H., Tao, M.-F., Ma, L., Wen, S.-Y., … Teng, Y.-C. (2016). RelB/NF-κB links cell cycle transition and apoptosis to endometrioid adenocarcinoma tumorigenesis. *Cell Death & Disease*, *7*(10), e2402–e2402. https://doi.org/10.1038/cddis.2016.309

Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, *32*(90001), 258D – 261. https://doi.org/10.1093/nar/gkh036

Genzel, Y., Ritter, J. B., König, S., Alt, R., & Reichl, U. (2005). Substitution of glutamine by pyruvate to reduce ammonia formation and growth inhibition of mammalian cells. *Biotechnology Progress*, *21*(1), 58–69. https://doi.org/10.1021/bp049827d

Ghaffari, N., Jardon, M. A., Krahn, N., Butler, M., Kennard, M., Turner, R. F. B., … Piret, J. M. (2019). Effects of cysteine, asparagine, or glutamine limitations in Chinese hamster ovary cell batch and fed-batch cultures. *Biotechnology Progress*, (December). https://doi.org/10.1002/btpr.2946

Giansanti, P., Tsiatsiani, L., Low, T. Y., & Heck, A. J. R. (2016). Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nature Protocols*, *11*(5), 993–1006. https://doi.org/10.1038/nprot.2016.057

Gingras, A. C., Raught, B., Gygi, S. P., Niedzwiecka, A., Miron, M., Burley, S. K., … Sonenberg, N. (2001). Hierarchical phosphorylation of the translation inhibitor 4E-BP1. *Genes and Development*, *15*(21), 2852–2864. https://doi.org/10.1101/gad.912401

Goshe, M. B. (2006). Characterizing phosphoproteins and phosphoproteomes using mass spectrometry. *Briefings in Functional Genomics and Proteomics*. https://doi.org/10.1093/bfgp/eli007

Groban, E. S., Narayanan, A., & Jacobson, M. P. (2006). Conformational changes in protein loops and helices induced by post-translational phosphorylation. *PLoS Computational Biology*, *2*(4), 238–250. https://doi.org/10.1371/journal.pcbi.0020032

Gross, S. R., & Kinzy, T. G. (2007). Improper Organization of the Actin Cytoskeleton Affects Protein Synthesis at Initiation. *Molecular and Cellular Biology*, *27*(5), 1974–1989. https://doi.org/10.1128/mcb.00832-06

Guntur, K. V. P., Guilherme, A., Xue, L., Chawla, A., & Czech, M. P. (2010). Map4k4 negatively regulates peroxisome proliferator-activated receptor (PPAR) gamma protein translation by suppressing the mammalian target of rapamycin (mTOR) signaling pathway in cultured adipocytes. *The Journal of Biological Chemistry*, *285*(9), 6595–6603. https://doi.org/10.1074/jbc.M109.068502

Gutierrez, J. M., & Lewis, N. E. (2015). Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. *Biotechnology Journal*. https://doi.org/10.1002/biot.201400647

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., & Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, *17*(10), 994–999. https://doi.org/10.1038/13690

Ha, T. K., & Lee, G. M. (2014). Effect of glutamine substitution by TCA cycle intermediates on the production and sialylation of Fc-fusion protein in Chinese hamster ovary cell culture. *Journal of Biotechnology*, *180*, 23–29. https://doi.org/10.1016/j.jbiotec.2014.04.002

Haas, A. L., & Siepmann, T. J. (1997). Pathways of ubiquitin conjugation. In *FASEB Journal* (Vol. 11, pp. 1257–1268). https://doi.org/10.1096/fasebj.11.14.9409544

Hacker, D. L., De Jesus, M., & Wurm, F. M. (2009). 25??years of recombinant proteins from reactor-grown cells - Where do we go from here? *Biotechnology Advances*. https://doi.org/10.1016/j.biotechadv.2009.05.008

Hammond, S., Swanberg, J. C., Kaplarevic, M., & Lee, K. H. (2011). Genomic sequencing and analysis of a Chinese hamster ovary cell line using Illumina sequencing technology. *BMC Genomics*,

12(1), 67. https://doi.org/10.1186/1471-2164-12-67

Han, S., Yu, B., Wang, Y., & Liu, Y. (2011). Role of plant autophagy in stress response. *Protein and Cell*, *2*(10), 784–791. https://doi.org/10.1007/s13238-011-1104-4

Han, Y. K., Ha, T. K., Lee, S. J., Lee, J. S., & Lee, G. M. (2011). Autophagy and apoptosis of recombinant Chinese hamster ovary cells during fed-batch culture: Effect of nutrient supplementation. *Biotechnology and Bioengineering*, *108*(9), 2182–2192. https://doi.org/10.1002/bit.23165

Han, Y., Liu, X. M., Liu, H., Li, S. C., Wu, B. C., Ye, L. L., … Chen, Z. L. (2006). Cultivation of Recombinant Chinese hamster ovary cells grown as suspended aggregates in stirred vessels. *Journal of Bioscience and Bioengineering*, *102*(5), 430–435. https://doi.org/10.1263/jbb.102.430

Hansen, H. G., Pristovšek, N., Kildegaard, H. F., & Lee, G. M. (2017). Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: Lessons learned and future directions. *Biotechnology Advances*. https://doi.org/10.1016/j.biotechadv.2016.11.008

Hara, K., Yonezawa, K., Kozlowski, M. T., Sugimoto, T., Andrabi, K., Weng, Q. P., … Avruch, J. (1997). Regulation of eIF-4E BP1 phosphorylation by mTOR. *Journal of Biological Chemistry*, *272*(42), 26457–26463. https://doi.org/10.1074/jbc.272.42.26457

Harton, J. A., Zika, E., & Ting, J. P. Y. (2001). The Histone Acetyltransferase Domains of CREB-binding Protein (CBP) and p300/CBP-associated Factor Are Not Necessary for Cooperativity with the Class II Transactivator. *Journal of Biological Chemistry*, *276*(42), 38715–38720. https://doi.org/10.1074/jbc.M106652200

Hauser, A., Penkert, M., & Hackenberger, C. P. R. (2017). Chemical Approaches to Investigate Labile Peptide and Protein Phosphorylation. *Accounts of Chemical Research*, *50*(8), 1883–1893. https://doi.org/10.1021/acs.accounts.7b00170

Hausmann, R., Chudobová, I., Spiegel, H., & Schillberg, S. (2018). Proteomic analysis of CHO cell lines producing high and low quantities of a recombinant antibody before and after selection with methotrexate. *Journal of Biotechnology*. https://doi.org/10.1016/j.jbiotec.2017.11.008

Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., … Lewis, N. E. (2016). A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Systems*, *3*(5), 434-443.e8. https://doi.org/10.1016/j.cels.2016.10.020

Henry, M., Coleman, O., Prashant, Clynes, M., & Meleady, P. (2017). Phosphopeptide enrichment and LC-MS/MS analysis to study the phosphoproteome of recombinant chinese hamster ovary cells. In *Methods in Molecular Biology* (Vol. 1603, pp. 195–208). https://doi.org/10.1007/978-1-4939-6972-2_13

Henry, M., Power, M., Kaushik, P., Coleman, O., Clynes, M., & Meleady, P. (2017a). Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift. *Journal of Proteome Research*, *16*(7), 2339–2358. https://doi.org/10.1021/acs.jproteome.6b00868

Henry, M., Power, M., Kaushik, P., Coleman, O., Clynes, M., & Meleady, P. (2017b). Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift. *Journal of Proteome Research*, *16*(7), 2339–2358. https://doi.org/10.1021/acs.jproteome.6b00868

Hernandez, I., Dhiman, H., Klanert, G., Jadhav, V., Auer, N., Hanscho, M., … Borth, N. (2019). Epigenetic regulation of gene expression in Chinese Hamster Ovary cells in response to the changing environment of a batch culture. *Biotechnology and Bioengineering*, *116*(3), 677–692. https://doi.org/10.1002/bit.26891

Hervouet, E., Vallette, F. M., & Cartron, P. F. (2009). Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. *Epigenetics*, *4*(7), 487–499. https://doi.org/10.4161/epi.4.7.9883

Higgs, R. E., Knierman, M. D., Gelfanova, V., Butler, J. P., & Hale, J. E. (2005). Comprehensive label-free method for the relative quantification of proteins from biological samples. *Journal of Proteome Research*, *4*(4), 1442–1450. https://doi.org/10.1021/pr050109b

Hinzke, T., Kouris, A., Hughes, R. A., Strous, M., & Kleiner, M. (2019). More is not always better:

Evaluation of 1D and 2D-LC-MS/MS methods for metaproteomics. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2019.00238

Hirasawa, N., Santini, F., & Beaven, M. A. (1995). Activation of the mitogen-activated protein kinase/cytosolic phospholipase A2pathway in a rat mast cell line: Indications of different pathways for release of arachidonic acid and secretory granules. *Journal of Immunology*, *154*(10), 5391–5402.

Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., … Tai, H. L. (2003). Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical Biochemist*, *24*(1), 3–12. https://doi.org/10.1146/annurev.bi.64.070195.001531

Ho, S. C. L., Koh, E. Y. C., Soo, B. P. C., Chao, S. H., & Yang, Y. (2016). Evaluating the use of a CpG free promoter for long-term recombinant protein expression stability in Chinese hamster ovary cells. *BMC Biotechnology*, *16*(1), 71. https://doi.org/10.1186/s12896-016-0300-y

Hong, J. K., Cho, S. M., & Yoon, S. K. (2010). Substitution of glutamine by glutamate enhances production and galactosylation of recombinant IgG in Chinese hamster ovary cells. *Applied Microbiology and Biotechnology*, *88*(4), 869–876. https://doi.org/10.1007/s00253-010-2790-1

Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., & Skrzypek, E. (2015). PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Research*, *43*(D1), D512–D520. https://doi.org/10.1093/nar/gku1267

Hrabák, J., Chudáčková, E., & Walková, R. (2013). Matrix-assisted laser desorption ionization-time of flight (maldi-tof) mass spectrometry for detection of antibiotic resistance mechanisms: from research to routine diagnosis. *Clinical Microbiology Reviews*, *26*(1), 103–114. https://doi.org/10.1128/CMR.00058-12

Hu, W., Berdugo, C., & Chalmers, J. J. (2011). The potential of hydrodynamic damage to animal cells of industrial relevance: Current understanding. *Cytotechnology*, *63*(5), 445–460. https://doi.org/10.1007/s10616-011-9368-3

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. https://doi.org/10.1038/nprot.2008.211

Huang, D. W., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., … Lempicki, R. A. (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Research*, *35*(suppl_2), W169–W175. https://doi.org/10.1093/nar/gkm415

Huang, Y.-M., Hu, W., Rustandi, E., Chang, K., Yusuf-Makagiansar, H., & Ryll, T. (2010). Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnology Progress*, *26*(5), 1400–1410. https://doi.org/10.1002/btpr.436

Humphrey, S. J., James, D. E., & Mann, M. (2015). Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. *Trends in Endocrinology and Metabolism*. https://doi.org/10.1016/j.tem.2015.09.013

Jardon, M. A., Sattha, B., Braasch, K., Leung, A. O., Côté, H. C. F., Butler, M., … Piret, J. M. (2012). Inhibition of glutamine-dependent autophagy increases t-PA production in CHO Cell fed-batch processes. *Biotechnology and Bioengineering*, *109*(5), 1228–1238. https://doi.org/10.1002/bit.24393

Jayapal, K. P., Wlaschin, K. F., Hu, W. S., & Yap, M. G. S. (2007). Recombinant protein therapeutics from CHO Cells - 20 years and counting. *Chemical Engineering Progress*, *103*(10), 40–47.

Jedrychowski, M. P., Huttlin, E. L., Haas, W., Sowa, M. E., Rad, R., & Gygi, S. P. (2011). Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Molecular & Cellular Proteomics : MCP*, *10*(12), M111.009910. https://doi.org/10.1074/mcp.M111.009910

Jiang, Z., Huang, Y., & Sharfstein, S. T. (2006). Regulation of recombinant monoclonal antibody production in chinese hamster ovary cells: a comparative study of gene copy number, mRNA level, and protein expression. *Biotechnology Progress*, *22*(1), 313–318.

https://doi.org/10.1021/bp0501524

Jiang, Z., & Sharfstein, S. T. (2008). Sodium butyrate stimulates monoclonal antibody over-expression in CHO cells by improving gene accessibility. *Biotechnology and Bioengineering*, *100*(1), 189–194. https://doi.org/10.1002/bit.21726

Jiang, Z., & Sharfstein, S. T. (2009). Characterization of gene localization and accessibility in DHFR-amplified CHO cells. *Biotechnology Progress*, *25*(1), 296–300.

Jing, X., Wang, T., Huang, S., Glorioso, J. C., & Albers, K. M. (2012). The transcription factor Sox11 promotes nerve regeneration through activation of the regeneration-associated gene Sprr1a. *Experimental Neurology*, *233*(1), 221–232. https://doi.org/10.1016/j.expneurol.2011.10.005

Johannessen, M., & Moens, U. (2007). Multisite phosphorylation of the cAMP response element-binding protein (CREB) by a diversity of protein kinases. *Frontiers in Bioscience*. https://doi.org/10.2741/2190

Josse, L., Xie, J., Proud, C. G., & Smales, C. M. (2016). mTORC1 signalling and eIF4E/4E-BP1 translation initiation factor stoichiometry influence recombinant protein productivity from GS-CHOK1 cells. *Biochemical Journal*, *473*(24), 4651–4664. https://doi.org/10.1042/BCJ20160845

Kaas, C. S., Kristensen, C., Betenbaugh, M. J., & Andersen, M. R. (2015). Sequencing the CHO DXB11 genome reveals regional variations in genomic stability and haploidy. *BMC Genomics*, *16*(1), 160. https://doi.org/10.1186/s12864-015-1391-x

Käll, L., & Vitek, O. (2011). Computational mass spectrometry-based proteomics. *PLoS Computational Biology*, *7*(12), 1–7. https://doi.org/10.1371/journal.pcbi.1002277

Kamada, Y., Yoshino, K. -i., Kondo, C., Kawamata, T., Oshiro, N., Yonezawa, K., & Ohsumi, Y. (2010). Tor Directly Controls the Atg1 Kinase Complex To Regulate Autophagy. *Molecular and Cellular Biology*, *30*(4), 1049–1058. https://doi.org/10.1128/MCB.01344-09

Karas, M., & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, *60*(20), 2299–2301. https://doi.org/10.1021/ac00171a028

Kast, D. J., & Dominguez, R. (2017). The Cytoskeleton–Autophagy Connection. *Current Biology*, *27*(8), R318–R326. https://doi.org/10.1016/j.cub.2017.02.061

Kaufmann, H., Mazur, X., Fussenegger, M., & Bailey, J. E. (1999). Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells. *Biotechnology and Bioengineering*, *63*(5), 573–582. https://doi.org/10.1002/(SICI)1097-0290(19990605)63:5<573::AID-BIT7>3.0.CO;2-Y

Kaushik, P., Henry, M., Clynes, M., & Meleady, P. (2018). The Expression Pattern of the Phosphoproteome Is Significantly Changed During the Growth Phases of Recombinant CHO Cell Culture. *Biotechnology Journal*. https://doi.org/10.1002/biot.201700221

Khan, A. S., Murray, M. J., Ho, C. M. K., Zuercher, W. J., Reeves, M. B., & Strang, B. L. (2017). High-throughput screening of a GlaxoSmithKline protein kinase inhibitor set identifies an inhibitor of human cytomegalovirus replication that prevents CREB and histone H3 post-translational modification. *The Journal of General Virology*, *98*(4), 754–768. https://doi.org/10.1099/jgv.0.000713

Kholodenko, B. N. (2002). MAP kinase cascade signaling and endocytic trafficking: A marriage of convenience? *Trends in Cell Biology*. https://doi.org/10.1016/S0962-8924(02)02251-1

Kildegaard, H. F., Baycin-Hizal, D., Lewis, N. E., & Betenbaugh, M. J. (2013). The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. *Current Opinion in Biotechnology*. https://doi.org/10.1016/j.copbio.2013.02.007

Kim, D. Y., Chaudhry, M. A., Kennard, M. L., Jardon, M. A., Braasch, K., Dionne, B., … Piret, J. M. (2013). Fed-batch CHO cell t-PA production and feed glutamine replacement to reduce ammonia production. *Biotechnology Progress*, *29*(1), 165–175. https://doi.org/10.1002/btpr.1658

Kim, J., Kundu, M., Viollet, B., & Guan, K. L. (2011). AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. *Nature Cell Biology*, *13*(2), 132–141. https://doi.org/10.1038/ncb2152

Kim, M. S., & Pandey, A. (2012). Electron transfer dissociation mass spectrometry in proteomics. *Proteomics*. https://doi.org/10.1002/pmic.201100517

Kim, S., & Coulombe, P. A. (2010). Emerging role for the cytoskeleton as an organizer and regulator of translation. *Nature Reviews Molecular Cell Biology*, *11*(1), 75–81. https://doi.org/10.1038/nrm2818

Kim, S. H., Trinh, A. T., Larsen, M. C., Mastrocola, A. S., Jefcoate, C. R., Bushel, P. R., & Tibbetts, R. S. (2016). Tunable regulation of CREB DNA binding activity couples genotoxic stress response and metabolism. *Nucleic Acids Research*, *44*(20), 9667–9680. https://doi.org/10.1093/nar/gkw643

Kim, Y. J., Baek, E., Lee, J. S., & Lee, G. M. (2013). Autophagy and its implication in Chinese hamster ovary cell culture. *Biotechnology Letters*. https://doi.org/10.1007/s10529-013-1276-5

King, C. a. (2013). Kaposi's sarcoma-associated herpesvirus kaposin B induces unique monophosphorylation of STAT3 at serine 727 and MK2-mediated inactivation of the STAT3 transcriptional repressor TRIM28. *Journal of Virology*, *87*(15), 8779–8791. https://doi.org/10.1128/JVI.02976-12

Knight, Z. A., Schilling, B., Row, R. H., Kenski, D. M., Gibson, B. W., & Shokat, K. M. (2003). Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nature Biotechnology*, *21*(9), 1047–1054. https://doi.org/10.1038/nbt863

Koedoot, E., Smid, M., Foekens, J. A., Martens, J. W. M., Le Dévédec, S. E., & van de Water, B. (2019). Co-regulated gene expression of splicing factors as drivers of cancer progression. *Scientific Reports*, *9*(1), 5484. https://doi.org/10.1038/s41598-019-40759-4

Köhler, G., & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, *256*(5517), 495–497.

Konermann, L., Ahadi, E., Rodriguez, A. D., & Vahidi, S. (2013). Unraveling the mechanism of electrospray ionization. *Analytical Chemistry*, *85*(1), 2–9. https://doi.org/10.1021/ac302789c

Kovács, K. A., Steinmann, M., Halfon, O., Magistretti, P. J., & Cardinaux, J. R. (2015). Complex regulation of CREB-binding protein by homeodomain-interacting protein kinase 2. *Cellular Signalling*, *27*(11), 2252–2260. https://doi.org/10.1016/j.cellsig.2015.08.001

Krämer, A., Green, J., Pollard Jr, J., & Tugendreich, S. (2013). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*, *30*(4), 523–530. https://doi.org/10.1093/bioinformatics/btt703

Krämer, A., Green, J., Pollard, J., & Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*, *30*(4), 523–530. https://doi.org/10.1093/bioinformatics/btt703

Kremkow, B. G., Baik, J. Y., MacDonald, M. L., & Lee, K. H. (2015). CHOgenome.org 2.0: Genome resources and website updates. *Biotechnology Journal*. https://doi.org/10.1002/biot.201400646

Kumar, N., Gammell, P., & Clynes, M. (2007). Proliferation control strategies to improve productivity and survival during CHO based production culture: A summary of recent methods employed and the effects of proliferation control in product secreting CHO cell lines. In *Cytotechnology*. https://doi.org/10.1007/s10616-007-9047-6

Kumar, N., Gammell, P., Meleady, P., Henry, M., & Clynes, M. (2008). Differential protein expression following low temperature culture of suspension CHO-K1 cells. *BMC Biotechnology*, *8*(1), 42. https://doi.org/10.1186/1472-6750-8-42

Kung, H.-N., Marks, J. R., & Chi, J.-T. (2011). Glutamine Synthetase Is a Genetic Determinant of Cell Type–Specific Glutamine Independence in Breast Epithelia. *PLoS Genetics*, *7*(8), e1002229. https://doi.org/10.1371/journal.pgen.1002229

Kuo, C.-C., Chiang, A. W., Shamie, I., Samoudi, M., Gutierrez, J. M., & Lewis, N. E. (2018). The emerging role of systems biology for engineering protein production in CHO cells. *Current Opinion in Biotechnology*, *51*, 64–69. https://doi.org/10.1016/j.copbio.2017.11.015

Kurochkina, N., & Guha, U. (2013). SH3 domains: Modules of protein-protein interactions. *Biophysical Reviews*. https://doi.org/10.1007/s12551-012-0081-z

Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., & Jørgensen, T. J. D. (2005). Highly

selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular and Cellular Proteomics*, *4*(7), 873–886. https://doi.org/10.1074/mcp.T500007-MCP200

Laux, H., Romand, S., Ritter, A., Oertli, M., Fornaro, M., Jostock, T., & Wilms, B. (2013). Generation of genetically engineered CHO cell lines to support the production of a difficult to express therapeutic protein. *BMC Proceedings*, *7*(S6), P1. https://doi.org/10.1186/1753-6561-7-S6-P1

Le Fourn, V., Girod, P.-A., Buceta, M., Regamey, A., & Mermod, N. (2014). CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. *Metabolic Engineering*, *21*, 91–102. https://doi.org/10.1016/j.ymben.2012.12.003

Le Gallic, L., Virgilio, L., Cohen, P., Biteau, B., & Mavrothalassitis, G. (2004). ERF Nuclear Shuttling, a Continuous Monitor of Erk Activity That Links It to Cell Cycle Progression. *Molecular and Cellular Biology*, *24*(3), 1206–1218. https://doi.org/10.1128/mcb.24.3.1206-1218.2004

Leal, P., Garcia, P., Sandoval, A., Buchegger, K., Weber, H., Tapia, O., & Roa, J. C. (2013). AKT/mTOR substrate p70S6k is frequently phosphorylated in gallbladder cancer tissue and cell lines. *OncoTargets and Therapy*, *6*, 1373–1384. https://doi.org/10.2147/OTT.S46897

Lee, J. S., Grav, L. M., Lewis, N. E., & Kildegaard, H. F. (2015). CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. *Biotechnology Journal*. https://doi.org/10.1002/biot.201500082

Leong, D. S. Z., Teo, B. K. H., Tan, J. G. L., Kamari, H., Yang, Y. S., Zhang, P., & Ng, S. K. (2018). Application of maltose as energy source in protein-free CHO-K1 culture to improve the production of recombinant monoclonal antibody. *Scientific Reports*, *8*(1), 4037. https://doi.org/10.1038/s41598-018-22490-8

Levine, A. A., Guan, Z., Barco, A., Xu, S., Kandel, E. R., & Schwartz, J. H. (2005). CREB-binding protein controls response to cocaine by acetylating histones at the fosB promoter in the mouse striatum. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(52), 19186 LP – 19191. https://doi.org/10.1073/pnas.0509735102

Lewis, J. K., Wei, J., & Siuzdak, G. (2000). Matrix-assisted Laser Desorption / Ionization Mass Spectrometry in Peptide and Protein Analysis. *Encyclopedia of Analytical Chemistry*, 5880–5894. https://doi.org/10.1002/9780470027318.a1621

Lewis, N. E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., … Palsson, B. O. (2013a). Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nature Biotechnology*, *31*(8), 759–765. https://doi.org/10.1038/nbt.2624

Lewis, N. E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., … Palsson, B. O. (2013b). Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nature Biotechnology*, *31*(8), 759–765. https://doi.org/10.1038/nbt.2624

Li, L., Tibiche, C., Fu, C., Kaneko, T., Moran, M. F., Schiller, M. R., … Wang, E. (2012). The human phosphotyrosine signaling network: Evolution and hotspots of hijacking in cancer. *Genome Research*, *22*(7), 1222–1230. https://doi.org/10.1101/gr.128819.111

Li, S., Cha, S. W., Heffner, K., Hizal, D. B., Bowen, M. A., Chaerkady, R., … Lewis, N. E. (2019). Proteogenomic Annotation of Chinese Hamsters Reveals Extensive Novel Translation Events and Endogenous Retroviral Elements. *Journal of Proteome Research*, *18*(6), 2433–2445. https://doi.org/10.1021/acs.jproteome.8b00935

Li, Yaojun, Luo, Y., Wu, S., Gao, Y., Liu, Y., & Zheng, D. (2009). Nucleic acids in protein samples interfere with phosphopeptide identification by immobilized-metal-ion affinity chromatography and mass spectrometry. *Molecular Biotechnology*, *43*(1), 59–66. https://doi.org/10.1007/s12033-009-9176-6

Li, Yun, Jones, J. W., M. C. Choi, H., Sarkar, C., Kane, M. A., Koh, E. Y., … Wu, J. (2019). cPLA2 activation contributes to lysosomal defects leading to impairment of autophagy after spinal cord injury. *Cell Death & Disease*, *10*(7), 531. https://doi.org/10.1038/s41419-019-1764-1

Liu, B. A. (2017). Characterizing SH2 domain specificity and network interactions using SPOT peptide arrays. In *Methods in Molecular Biology* (Vol. 1555, pp. 357–373). https://doi.org/10.1007/978-1-4939-6762-9_20

Lombardi, B., Rendell, N., Edwards, M., Katan, M., & Zimmermann, J. G. (2015). Evaluation of phosphopeptide enrichment strategies for quantitative TMT analysis of complex network dynamics in cancer-associated cell signalling. *EuPA Open Proteomics*, *6*, 10–15. https://doi.org/10.1016/j.euprot.2015.01.002

Lopes, C. T., Franz, M., Kazi, F., Donaldson, S. L., Morris, Q., Bader, G. D., & Dopazo, J. (2011). Cytoscape Web: An interactive web-based network browser. In *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btq430

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lucas, B. K., Giere, L. M., DeMarco, R. A., Shen, A., Chisholm, V., & Crowley, C. W. (1996). High-level production of recombinant proteins in CHO cells using a dicistronic DHFR intron expression vector. *Nucleic Acids Research*, *24*(9), 1774–1779. https://doi.org/10.1093/nar/24.9.1774

Luche, S., Santoni, V., & Rabilloud, T. (2003). Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics*, *3*(3), 249–253. https://doi.org/10.1002/pmic.200390037

Lund, A. M., Kaas, C. S., Brandl, J., Pedersen, L. E., Kildegaard, H. F., Kristensen, C., & Andersen, M. R. (2017). Network reconstruction of the mouse secretory pathway applied on CHO cell transcriptome data. *BMC Systems Biology*, *11*(1), 37. https://doi.org/10.1186/s12918-017-0414-4

Lundgren, D. H., Hwang, S.-I., Wu, L., & Han, D. K. (2010). Role of spectral counting in quantitative proteomics. *Expert Review of Proteomics*, *7*(1), 39–53. https://doi.org/10.1586/epr.09.69

Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., & Horning, S. (2006). Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical Chemistry*, *78*(7), 2113–2120. https://doi.org/10.1021/ac0518811

Mamyrin, B. A. (2001). Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry*, *206*(3), 251–266. https://doi.org/10.1016/S1387-3806(00)00392-4

Mann, M., & Kelleher, N. L. (2008). Precision proteomics: the case for high resolution and high mass accuracy. *Proc Natl Acad Sci U S A*, *105*(47), 18132–18138. https://doi.org/10.1073/pnas.0800788105

Marginean, I., Nemes, P., Parvin, L., & Vertes, A. (2006). How much charge is there on a pulsating Taylor cone? *Applied Physics Letters*, *89*(6), 064104. https://doi.org/10.1063/1.2266889

Mariño, K., Bones, J., Kattla, J. J., & Rudd, P. M. (2010). A systematic approach to protein glycosylation analysis: A path through the maze. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.437

Masterton, R. J., Roobol, A., Al-Fageeh, M. B., Carden, M. J., & Smales, C. M. (2010). Post-translational events of a model reporter protein proceed with higher fidelity and accuracy upon mild hypothermic culturing of Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *105*(1), 215–220. https://doi.org/10.1002/bit.22533

Mattson, M. P., & Meffert, M. K. (2006). Roles for NF-κB in nerve cell survival, plasticity, and disease. *Cell Death and Differentiation*, *13*(5), 852–860. https://doi.org/10.1038/sj.cdd.4401837

Meleady, P. (2007). Proteomic profiling of recombinant cells from large-scale mammalian cell culture processes. In *Cytotechnology* (Vol. 53, pp. 23–31). https://doi.org/10.1007/s10616-007-9052-9

Meleady, P., Hoffrogge, R., Henry, M., Rupp, O., Bort, J. H., Clarke, C., … Borth, N. (2012). Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics. *Biotechnology and Bioengineering*, *109*(6), 1386–1394. https://doi.org/10.1002/bit.24476

Michalski, A., Cox, J., & Mann, M. (2011). More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *Journal of Proteome Research*, *10*(4), 1785–1793. https://doi.org/10.1021/pr101060v

Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. *BioEssays*, *35*(12), 1050–1055. https://doi.org/10.1002/bies.201300066

Mincheva, S., Garcera, A., Gou-Fabregas, M., Encinas, M., Dolcet, X., & Soler, R. M. (2011). The canonical nuclear factor-κB pathway regulates cell survival in a developmental model of spinal cord motoneurons. *Journal of Neuroscience*, *31*(17), 6493–6503. https://doi.org/10.1523/JNEUROSCI.0206-11.2011

Moschella, P. C., Rao, V. U., McDermott, P. J., & Kuppuswamy, D. (2007). Regulation of mTOR and S6K1 activation by the nPKC isoforms, PKCε and PKCδ, in adult cardiac muscle cells. *Journal of Molecular and Cellular Cardiology*, *43*(6), 754–766. https://doi.org/10.1016/j.yjmcc.2007.09.015

Murillo, J. R., Kuras, M., Rezeli, M., Milliotis, T., Betancourt, L., & Marko-Varga, G. (2018). Automated phosphopeptide enrichment from minute quantities of frozen malignant melanoma tissue. *PLOS ONE*, *13*(12), e0208562. https://doi.org/10.1371/journal.pone.0208562

Nasseri, S. S., Ghaffari, N., Braasch, K., Jardon, M. A., Butler, M., Kennard, M., … Piret, J. M. (2014). Increased CHO cell fed-batch monoclonal antibody production using the autophagy inhibitor 3-MA or gradually increasing osmolality. *Biochemical Engineering Journal*, *91*, 37–45. https://doi.org/10.1016/j.bej.2014.06.027

Nawrocki, J., Dunlap, C., Li, J., Zhao, J., McNeff, C. V., McCormick, A., & Carr, P. W. (2004). Part II. Chromatography using ultra-stable metal oxide-based stationary phases for HPLC. *Journal of Chromatography A*. https://doi.org/10.1016/j.chroma.2003.11.050

Nika, H., Nieves, E., Hawke, D. H., & Angeletti, R. H. (2013). Phosphopeptide enrichment by covalent chromatography after derivatization of protein digests immobilized on reversed-phase supports. *Journal of Biomolecular Techniques*, *24*(3), 154–177. https://doi.org/10.7171/jbt.13-2403-004

Nikolov, M., Schmidt, C., & Urlaub, H. (2012). Quantitative mass spectrometry-based proteomics: An overview. *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-61779-885-6_7

Noh, S. M., Shin, S., & Lee, G. M. (2018). Comprehensive characterization of glutamine synthetase-mediated selection for the establishment of recombinant CHO cells producing monoclonal antibodies. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-23720-9

Ochoa, D., Jarnuczak, A. F., Viéitez, C., Gehre, M., Soucheray, M., Mateus, A., … Beltrao, P. (2019). The functional landscape of the human phosphoproteome. *Nature Biotechnology*. https://doi.org/10.1038/s41587-019-0344-3

Oliveros, J. C. (2007). VENNY. An interactive tool for comparing lists with Venn Diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html. Retrieved from http://bioinfogp.cnnb.csic.es/tools/venny/index.html

Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., & Mann, M. (2006). Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell*, *127*(3), 635–648. https://doi.org/10.1016/j.cell.2006.09.026

Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., & Mann, M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods*, *4*(9), 709–712. https://doi.org/10.1038/nmeth1060

Oltean, S., & Bates, D. O. (2014). Hallmarks of alternative splicing in cancer. *Oncogene*, *33*(46), 5311–5318. https://doi.org/10.1038/onc.2013.533

Omasa, T., Onitsuka, M., & Kim, W.-D. (2010). Cell Engineering and Cultivation of Chinese Hamster Ovary (CHO) Cells. *Current Pharmaceutical Biotechnology*, *11*(3), 233–240. https://doi.org/10.2174/138920110791111960

Ong, S.-E., & Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology*, *1*(5), 252–262. https://doi.org/10.1038/nchembio736

Ortega-Martínez, S. (2015). A new perspective on the role of the CREB family of transcription factors in memory consolidation via adult hippocampal neurogenesis. *Frontiers in Molecular Neuroscience*, *8*(AUGUST), 46. https://doi.org/10.3389/fnmol.2015.00046

Park, J. H., Jin, J. H., Ji, I. J., An, H. J., Kim, J. W., & Lee, G. M. (2017). Proteomic analysis of host

cell protein dynamics in the supernatant of Fc-fusion protein-producing CHO DG44 and DUKX-B11 cell lines in batch and fed-batch cultures. *Biotechnology and Bioengineering*, *114*(10), 2267–2278. https://doi.org/10.1002/bit.26360

Paulo, J. A., Navarrete-Perea, J., Erickson, A. R., Knott, J., & Gygi, S. P. (2018). An Internal Standard for Assessing Phosphopeptide Recovery from Metal Ion/Oxide Enrichment Strategies. *Journal of the American Society for Mass Spectrometry*. https://doi.org/10.1007/s13361-018-1946-6

Pegram, L. M., Liddle, J. C., Xiao, Y., Hoh, M., Rudolph, J., Iverson, D. B., … Ahn, N. G. (2019). Activation loop dynamics are controlled by conformation-selective inhibitors of ERK2. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(31), 15463–15468. https://doi.org/10.1073/pnas.1906824116

Pereira, S., Kildegaard, H. F., & Andersen, M. R. (2018a). Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients. *Biotechnology Journal*, *13*(3), 1700499. https://doi.org/10.1002/biot.201700499

Pereira, S., Kildegaard, H. F., & Andersen, M. R. (2018b). Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients. *Biotechnology Journal*, *13*(3), 1700499. https://doi.org/10.1002/biot.201700499

Perkins, D N, Pappin, D. J., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, *20*(18), 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

Perkins, David N., Pappin, D. J. C., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. In *Electrophoresis* (Vol. 20, pp. 3551–3567). https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

Peroutka, R. J., Elshourbagy, N., Piech, T., & Butt, T. R. (2008). Enhanced protein expression in mammalian cells using engineered SUMO fusions: secreted phospholipase A2. *Protein Science : A Publication of the Protein Society*. https://doi.org/10.1110/ps.035576.108.of

Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., … Aebersold, R. (2010). High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nature Methods*, *7*(1), 43–46. https://doi.org/10.1038/nmeth.1408

Popovic, D., Vucic, D., & Dikic, I. (2014). Ubiquitination in disease pathogenesis and treatment. *Nature Medicine*. https://doi.org/10.1038/nm.3739

Potel, C. M., Lemeer, S., & Heck, A. J. R. (2019). Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Analytical Chemistry*. https://doi.org/10.1021/acs.analchem.8b04746

Prabakaran, S., Lippens, G., Steen, H., & Gunawardena, J. (2012). Post-translational modification: Nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. https://doi.org/10.1002/wsbm.1185

PUCK, T. T., CIECIURA, S. J., & ROBINSON, A. (1958). Genetics of somatic mammalian cells. III. Long-term cultivation of euploid cells from human and animal subjects. *The Journal of Experimental Medicine*, *108*(6), 945–956. https://doi.org/10.1084/jem.108.6.945

Puente, C., Hendrickson, R. C., & Jiang, X. (2016). Nutrient-regulated phosphorylation of ATG13 inhibits starvation-induced autophagy. *Journal of Biological Chemistry*, *291*(11), 6026–6035. https://doi.org/10.1074/jbc.M115.689646

Qin, J. Y., Zhang, L., Clift, K. L., Hulur, I., Xiang, A. P., Ren, B.-Z., & Lahn, B. T. (2010). Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PloS One*, *5*(5), e10611. https://doi.org/10.1371/journal.pone.0010611

Qin, X., Jiang, B., & Zhang, Y. (2016). 4E-BP1, a multifactor regulated multifunctional protein. *Cell Cycle*. https://doi.org/10.1080/15384101.2016.1151581

R Core Team. (2014). R Core Team (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. URL Http://Www.R-Project.Org/.*

Rajendra, Y., Kiseljak, D., Baldi, L., Hacker, D. L., & Wurm, F. M. (2012). Reduced glutamine concentration improves protein production in growth-arrested CHO-DG44 and HEK-293E cells. *Biotechnology Letters*, *34*(4), 619–626. https://doi.org/10.1007/s10529-011-0809-z

Ramroop, J. R., Stein, M. N., & Drake, J. M. (2018a). Impact of Phosphoproteomics in the Era of Precision Medicine for Prostate Cancer. *Frontiers in Oncology*, *8*(FEB). https://doi.org/10.3389/fonc.2018.00028

Ramroop, J. R., Stein, M. N., & Drake, J. M. (2018b). Impact of Phosphoproteomics in the Era of Precision Medicine for Prostate Cancer. *Frontiers in Oncology*, *8*(FEB), 28. https://doi.org/10.3389/fonc.2018.00028

Rao, X., Huang, X., Zhou, Z., & Lin, X. (2013). An improvement of the 2ˆ(–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostatistics, Bioinformatics and Biomathematics*, *3*(3), 71–85.

Reinhart, D., Damjanovic, L., Castan, A., Ernst, W., & Kunert, R. (2018a). Differential gene expression of a feed-spiked super-producing CHO cell line. *Journal of Biotechnology*, *285*, 23–37. https://doi.org/10.1016/j.jbiotec.2018.08.013

Reinhart, D., Damjanovic, L., Castan, A., Ernst, W., & Kunert, R. (2018b). Differential gene expression of a feed-spiked super-producing CHO cell line. *Journal of Biotechnology*, *285*, 23–37. https://doi.org/10.1016/j.jbiotec.2018.08.013

Resemann, A., Wunderlich, D., Rothbauer, U., Warscheid, B., Leonhardt, H., Fuchser, J., … Suckau, D. (2010). Top-down de novo protein sequencing of a 13.6 kDa camelid single heavy chain antibody by matrix-assisted laser desorption ionization-time-of-flight/ time-of-flight mass spectrometry. *Analytical Chemistry*, *82*(8), 3283–3292. https://doi.org/10.1021/ac1000515

Riggs, L., Seeley, E. H., & Regnier, F. E. (2005). Quantification of phosphoproteins with global internal standard technology. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *817*(1), 89–96. https://doi.org/10.1016/j.jchromb.2004.04.037

Riley, N. M., & Coon, J. J. (2018). The Role of Electron Transfer Dissociation in Modern Proteomics. *Analytical Chemistry*, *90*(1), 40–64. https://doi.org/10.1021/acs.analchem.7b04810

Roobol, A., Carden, M. J., Newsam, R. J., & Smales, C. M. (2009). Biochemical insights into the mechanisms central to the response of mammalian cells to cold stress and subsequent rewarming. *FEBS Journal*, *276*(1), 286–302. https://doi.org/10.1111/j.1742-4658.2008.06781.x

Roobol, A., Roobol, J., Carden, M. J., Bastide, A., Willis, A. E., Dunn, W. B., … Smales, C. M. (2011). ATR (ataxia telangiectasia mutated-and Rad3-related kinase) is activated by mild hypothermia in mammalian cells and subsequently activates p53. *Biochem. J*, *435*, 499–508. https://doi.org/10.1042/BJ20101303

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., … Pappin, D. J. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics : MCP*, *3*(12), 1154–1169. https://doi.org/10.1074/mcp.M400129-MCP200

Roux, A., Lison, D., Junot, C., & Heilier, J. F. (2011). Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review. *Clinical Biochemistry*. https://doi.org/10.1016/j.clinbiochem.2010.08.016

Sakamoto, K., Huang, B.-W., Iwasaki, K., Hailemariam, K., Ninomiya-Tsuji, J., & Tsuji, Y. (2010). Regulation of Genotoxic Stress Response by Homeodomain-interacting Protein Kinase 2 through Phosphorylation of Cyclic AMP Response Element-binding Protein at Serine 271. *Molecular Biology of the Cell*. https://doi.org/10.1091/mbc.e10-01-0015

Salvi, G., De Los Rios, P., & Vendruscolo, M. (2005). Effective interactions between chaotropic agents and proteins. *Proteins: Structure, Function and Genetics*, *61*(3), 492–499. https://doi.org/10.1002/prot.20626

Samson, T., Welch, C., Monaghan-Benson, E., Hahn, K. M., & Burridge, K. (2010). Endogenous RhoG is rapidly activated after epidermal growth factor stimulation through multiple guanine-nucleotide exchange factors. *Molecular Biology of the Cell*, *21*(9), 1629–1642. https://doi.org/10.1091/mbc.E09-09-0809

Savitski, M. M., Scholten, A., Sweetman, G., Mathieson, T., & Bantscheff, M. (2010). Evaluation of data analysis strategies for improved mass spectrometry-based phosphoproteomics. *Analytical Chemistry*, *82*(23), 9843–9849. https://doi.org/10.1021/ac102083q

Schelletter, L., Albaum, S., Walter, S., Noll, T., & Hoffrogge, R. (2019). Clonal variations in CHO IGF signaling investigated by SILAC-based phosphoproteomics and LFQ-MS. *Applied Microbiology and Biotechnology*, *103*(19), 8127–8143. https://doi.org/10.1007/s00253-019-10020-z

Schug, J. (2008). Using TESS to predict transcription factor binding sites in DNA sequence. *Current Protocols in Bioinformatics*, *Chapter 2*(SUPPL. 21), Unit 2.6. https://doi.org/10.1002/0471250953.bi0206s21

Sgouras, D. N., Athanasiou, M. A., Beal, G. J., Fisher, R. J., Blair, D. G., & Mavrothalassitis, G. J. (1995). ERF: an ETS domain protein with strong transcriptional repressor activity, can suppress ets-associated tumorigenesis and is regulated by phosphorylation during cell cycle and mitogenic stimulation. *The EMBO Journal*, *14*(19), 4781–4793. https://doi.org/10.1002/j.1460-2075.1995.tb00160.x

Sha, S., Huang, Z., Wang, Z., & Yoon, S. (2018). Mechanistic modeling and applications for CHO cell culture development and production. *Current Opinion in Chemical Engineering*. https://doi.org/10.1016/j.coche.2018.08.010

Shah, O. J., Anthony, J. C., Kimball, S. R., & Jefferson, L. S. (2000). 4E-BP1 and S6K1: translational integration sites for nutritional and hormonal information in muscle. *American Journal of Physiology. Endocrinology and Metabolism*, *279*(4), E715-29.

Shanware, N. P., Zhan, L., Hutchinson, J. A., Kim, S. H., Williams, L. M., & Tibbetts, R. S. (2010). Conserved and Distinct Modes of CREB/ATF Transcription Factor Regulation by PP2A/B56γ and Genotoxic Stress. *PLoS ONE*, *5*(8), e12173. https://doi.org/10.1371/journal.pone.0012173

Sharma, K., D'Souza, R. C. J., Tyanova, S., Schaab, C., Wi??niewski, J., Cox, J., & Mann, M. (2014). Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Reports*, *8*(5), 1583–1594. https://doi.org/10.1016/j.celrep.2014.07.036

Shin, Y. J., Kim, Y. B., & Kim, J. H. (2013). Protein kinase CK2 phosphorylates and activates p21-activated kinase 1. *Molecular Biology of the Cell*, *24*(18), 2990–2999. https://doi.org/10.1091/mbc.E13-04-0204

Singh, V., Ram, M., Kumar, R., Prasad, R., Roy, B. K., & Singh, K. K. (2017). Phosphorylation: Implications in Cancer. *The Protein Journal*, *36*(1), 1–6. https://doi.org/10.1007/s10930-017-9696-z

Smith, R. J. (1990). Glutamine Metabolism and Its Physiologic Importance. *Journal of Parenteral and Enteral Nutrition*, *14*(4_suppl), 40S-44S. https://doi.org/10.1177/014860719001400402

Solari, F. A., Dell'Aica, M., Sickmann, A., & Zahedi, R. P. (2015). Why phosphoproteomics is still a challenge. *Molecular BioSystems*, *11*(6), 1487–1493. https://doi.org/10.1039/c5mb00024f

Spiro, R. G. (2002). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, *12*(4), 43R-56R. https://doi.org/10.1093/glycob/12.4.43R

Steinert, P. M., & Marekov, L. N. (1997). Direct evidence that involucrin is a major early isopeptide cross- linked component of the keratinocyte cornified cell envelope. *Journal of Biological Chemistry*, *272*(3), 2021–2030. https://doi.org/10.1074/jbc.272.3.2021

Stolfa, G., Smonskey, M. T., Boniface, R., Hachmann, A.-B., Gulde, P., Joshi, A. D., … Campbell, A. (2017). CHO-Omics Review: The Impact of Current and Emerging Technologies on Chinese Hamster Ovary Based Bioproduction. *Biotechnology Journal*, 1700227. https://doi.org/10.1002/biot.201700227

Stultz, C. M., Levin, A. D., & Edelman, E. R. (2002). Phosphorylation-induced conformational changes in a mitogen-activated protein kinase substrate: Implications for tyrosine hydroxylase activation. *Journal of Biological Chemistry*, *277*(49), 47653–47661. https://doi.org/10.1074/jbc.M208755200

Sun, O. H., & Lee, G. M. (2008). Nutrient deprivation induces autophagy as well as apoptosis in Chinese hamster ovary cell culture. *Biotechnology and Bioengineering*, *99*(3), 678–685.

https://doi.org/10.1002/bit.21589

Sun, W., Wu, S., Wang, X., Zheng, D., & Gao, Y. (2004). A systematical analysis of tryptic peptide identification with reverse phase liquid chromatography and electrospray ion trap mass spectrometry. *Genomics, Proteomics & Bioinformatics / Beijing Genomics Institute*, *2*(3), 174–183. https://doi.org/10.1016/S1672-0229(04)02023-6

Sun, Y., Liu, W. Z., Liu, T., Feng, X., Yang, N., & Zhou, H. F. (2015). Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. *Journal of Receptors and Signal Transduction*, *35*(6), 600–604. https://doi.org/10.3109/10799893.2015.1030412

Svensmark, J. H., & Brakebusch, C. (2019). Rho GTPases in cancer: friend or foe? *Oncogene*. https://doi.org/10.1038/s41388-019-0963-7

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., … Von Mering, C. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Tan, H.-Y., Wang, N., Man, K., Tsao, S.-W., Che, C.-M., & Feng, Y. (2015). Autophagy-induced RelB/p52 activation mediates tumour-associated macrophage repolarisation and suppression of hepatocellular carcinoma by natural compound baicalin. *Cell Death & Disease*, *6*(10), e1942–e1942. https://doi.org/10.1038/cddis.2015.271

Tan, J. G. L., Lee, Y. Y., Wang, T., Yap, M. G. S., Tan, T. W., & Ng, S. K. (2015). Heat shock protein 27 overexpression in CHO cells modulates apoptosis pathways and delays activation of caspases to improve recombinant monoclonal antibody titre in fed-batch bioreactors. *Biotechnology Journal*, *10*(5), 790–800. https://doi.org/10.1002/biot.201400764

Tan, V. M., Cheng, L. C., & Drake, J. M. (2016). Complementing genomics and transcriptomics: Phosphoproteomics illuminates systems biology in prostate cancer. *Molecular & Cellular Oncology*, *3*(6), e1246075. https://doi.org/10.1080/23723556.2016.1246075

Tanel, A., & Averill-Bates, D. A. (2007). P38 and ERK mitogen-activated protein kinases mediate acrolein-induced apoptosis in Chinese hamster ovary cells. *Cellular Signalling*, *19*(5), 968–977. https://doi.org/10.1016/j.cellsig.2006.10.014

Tape, C. J., Worboys, J. D., Sinclair, J., Gourlay, R., Vogt, J., McMahon, K. M., … Jørgensen, C. (2014). Reproducible automated phosphopeptide enrichment using magnetic TiO2 and Ti-IMAC. *Analytical Chemistry*, *86*(20), 10296–10302. https://doi.org/10.1021/ac5025842

Taschwer, M., Hackl, M., Hernández Bort, J. A., Leitner, C., Kumar, N., Puc, U., … Borth, N. (2012). Growth, productivity and protein glycosylation in a CHO EpoFc producer cell line adapted to glutamine-free growth. *Journal of Biotechnology*, *157*(2), 295–303. https://doi.org/10.1016/j.jbiotec.2011.11.014

Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., & Mechtler, K. (2011). Universal and confident phosphorylation site localization using phosphoRS. *Journal of Proteome Research*, *10*(12), 5354–5362. https://doi.org/10.1021/pr200611n

Tayi, V. S., & Butler, M. (2015). Isolation and quantification of N-glycans from immunoglobulin G antibodies for quantitative glycosylation analysis. *Journal of Biological Methods*, *2*(2), 19. https://doi.org/10.14440/jbm.2015.52

Tejwani, V., Andersen, M. R., Nam, J. H., & Sharfstein, S. T. (2018). Glycoengineering in CHO Cells: Advances in Systems Biology. *Biotechnology Journal*, *13*(3), 1700234. https://doi.org/10.1002/biot.201700234

Thingholm, T. E., Jensen, O. N., & Larsen, M. R. (2009). Enrichment and Separation of Mono- and Multiply Phosphorylated Peptides Using Sequential Elution from IMAC Prior to Mass Spectrometric Analysis. In *Methods in molecular biology (Clifton, N.J.)* (pp. 67–78). https://doi.org/10.1007/978-1-60327-834-8_6

Thingholm, T. E., & Larsen, M. R. (2016a). Phosphopeptide Enrichment by Immobilized Metal Affinity Chromatography (pp. 123–133). https://doi.org/10.1007/978-1-4939-3049-4_8

Thingholm, T. E., & Larsen, M. R. (2016b). Sequential elution from IMAC (SIMAC): An efficient

method for enrichment and separation of mono-and multi-phosphorylated peptides. In *Methods in Molecular Biology* (Vol. 1355, pp. 147–160). https://doi.org/10.1007/978-1-4939-3049-4_10

Thingholm, T. E., & Larsen, M. R. (2016c). The use of titanium dioxide for selective enrichment of phosphorylated peptides. In *Methods in Molecular Biology* (Vol. 1355, pp. 135–146). https://doi.org/10.1007/978-1-4939-3049-4_9

Tong, K., Wang, Y., & Su, Z. (2017). Phosphotyrosine signalling and the origin of animal multicellularity. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1860), 20170681. https://doi.org/10.1098/rspb.2017.0681

Torres, M., Berrios, J., Rigual, Y., Latorre, Y., Vergara, M., Dickson, A. J., & Altamirano, C. (2019). Metabolic flux analysis during galactose and lactate co-consumption reveals enhanced energy metabolism in continuous CHO cell cultures. *Chemical Engineering Science*, *205*, 201–211. https://doi.org/10.1016/j.ces.2019.04.049

Trinh, A. T., Kim, S. H., Chang, H. Y., Mastrocola, A. S., & Tibbetts, R. S. (2013). Cyclin-dependent kinase 1-dependent phosphorylation of cAMP response element-binding protein decreases chromatin occupancy. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M113.464057

Ubersax, J. a, & Ferrell, J. E. (2007). Mechanisms of specificity in protein phosphorylation. *Nature Reviews. Molecular Cell Biology*, *8*(7), 530–541. https://doi.org/10.1038/nrm2203

Underhill, M. F., Birch, J. R., Smales, C. M., & Naylor, L. H. (2005). eIF2α phosphorylation, stress perception, and the shutdown of global protein synthesis in cultured CHO cells. *Biotechnology and Bioengineering*, *89*(7), 805–814. https://doi.org/10.1002/bit.20403

Urquhart, L. (2018). Market watch: Top drugs and companies by sales in 2017. *Nature Reviews Drug Discovery*. https://doi.org/10.1038/nrd.2018.42

Veggiani, G., Gerpe, M. C. R., Sidhu, S. S., & Zhang, W. (2019). Emerging drug development technologies targeting ubiquitination for cancer therapeutics. *Pharmacology & Therapeutics*, *199*, 139–154. https://doi.org/10.1016/j.pharmthera.2019.03.003

Veith, N., Ziehr, H., MacLeod, R. A. F., & Reamon-Buettner, S. M. (2016). Mechanisms underlying epigenetic and transcriptional heterogeneity in Chinese hamster ovary (CHO) cell lines. *BMC Biotechnology*, *16*(1), 6. https://doi.org/10.1186/s12896-016-0238-0

Villén, J., & Gygi, S. P. (2008). The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protocols*, *3*(10), 1630–1638. https://doi.org/10.1038/nprot.2008.150

Vogel, W. K., Gafken, P. R., Leid, M., & Filtz, T. M. (2014). Kinetic analysis of BCL11B multisite phosphorylation-dephosphorylation and coupled sumoylation in primary thymocytes by multiple reaction monitoring mass spectroscopy. *Journal of Proteome Research*, *13*(12), 5860–5868. https://doi.org/10.1021/pr5007697

Vyse, S., Desmond, H., & Huang, P. H. (2017). Advances in mass spectrometry based strategies to study receptor tyrosine kinases. *IUCrJ*. https://doi.org/10.1107/S2052252516020546

Wahrheit, J., Nicolae, A., & Heinzle, E. (2014). Dynamics of growth and metabolism controlled by glutamine availability in Chinese hamster ovary cells. *Applied Microbiology and Biotechnology*, *98*(4), 1771–1783. https://doi.org/10.1007/s00253-013-5452-2

Walsh, G. (2014). Biopharmaceutical benchmarks 2014. *Nature Biotechnology*, *32*(10), 992–1000. https://doi.org/10.1038/nbt.3040

Walsh, G. (2018). Biopharmaceutical benchmarks 2018. *Nature Biotechnology*, *36*(12), 1136–1145. https://doi.org/10.1038/nbt.4305

Walther, T. C., & Mann, M. (2010). Mass spectrometry-based proteomics in cell biology. *The Journal of Cell Biology*, *190*(4), 491–500. https://doi.org/10.1083/jcb.201004052

Wang, H., Xu, J., Lazarovici, P., Quirion, R., & Zheng, W. (2018). cAMP Response Element-Binding Protein (CREB): A Possible Signaling Molecule Link in the Pathophysiology of Schizophrenia. *Frontiers in Molecular Neuroscience*. https://doi.org/10.3389/fnmol.2018.00255

Wang, J., Mauvoisin, D., Martin, E., Atger, F., Galindo, A. N., Dayon, L., … Gachon, F. (2017).

Nuclear Proteomics Uncovers Diurnal Regulatory Landscapes in Mouse Liver. *Cell Metabolism*. https://doi.org/10.1016/j.cmet.2016.10.003

Wang, Z. G., Lv, N., Bi, W. Z., Zhang, J. L., & Ni, J. Z. (2015). Development of the affinity materials for phosphorylated proteins/peptides enrichment in phosphoproteomics analysis. *ACS Applied Materials and Interfaces*, *7*(16), 8377–8392. https://doi.org/10.1021/acsami.5b01254

Wickham, H. (2009). *ggplot2*. *ggplot2*. https://doi.org/10.1007/978-0-387-98141-3

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., … Hochstrasser, D. F. (1996). From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/Technology*, *14*(1), 61–65. https://doi.org/10.1038/nbt0196-61

Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, *6*(5), 359–362. https://doi.org/10.1038/nmeth.1322

Wiza, C., Nascimento, E. B. M., & Ouwens, D. M. (2012). Role of PRAS40 in Akt and mTOR signaling in health and disease. *American Journal of Physiology-Endocrinology and Metabolism*, *302*(12), E1453–E1460. https://doi.org/10.1152/ajpendo.00660.2011

Wright, J. C., Collins, M. O., Yu, L., Käll, L., Brosch, M., & Choudhary, J. S. (2012). Enhanced Peptide Identification by Electron Transfer Dissociation Using an Improved Mascot Percolator. *Molecular & Cellular Proteomics*, *11*(8), 478–491. https://doi.org/10.1074/mcp.O111.014522

Wu, D., Asiedu, M., Adelstein, R. S., & Wei, Q. (2006). A novel guanine nucleotide exchange factor MyoGEF is required for cytokinesis. *Cell Cycle*, *5*(11), 1234–1239. https://doi.org/10.4161/cc.5.11.2815

Wu, X., & McMurray, C. T. (2001). Calmodulin kinase II attenuation of gene transcription by preventing cAMP response element-binding protein (CREB) dimerization and binding of the creb-binding protein. *Journal of Biological Chemistry*, *276*(3), 1735–1741. https://doi.org/10.1074/jbc.M006727200

Wurm, F. M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology*, *22*(11), 1393–1398. https://doi.org/10.1038/nbt1026

Wurm, F. M., & Hacker, D. (2011). First CHO genome. *Nature Biotechnology*, *29*(8), 718–720. https://doi.org/10.1038/nbt.1943

Xu, N., Ma, C., Ou, J., Sun, W. W., Zhou, L., Hu, H., & Liu, X. M. (2017). Comparative proteomic analysis of three Chinese hamster ovary (CHO) host cells. *Biochemical Engineering Journal*, *124*, 122–129. https://doi.org/10.1016/j.bej.2017.05.007

Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., … Wang, J. (2011a). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nature Biotechnology*, *29*(8), 735–741. https://doi.org/10.1038/nbt.1932

Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., … Wang, J. (2011b). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nature Biotechnology*, *29*(8), 735–741. https://doi.org/10.1038/nbt.1932

Yamada, S., Kamata, T., Nawa, H., Sekijima, T., & Takei, N. (2019). AMPK activation, eEF2 inactivation, and reduced protein synthesis in the cerebral cortex of hibernating chipmunks. *Scientific Reports*, *9*(1), 11904. https://doi.org/10.1038/s41598-019-48172-7

Yan, G. R., & He, Q. Y. (2008). Functional proteomics to identify critical proteins in signal transduction pathways. *Amino Acids*. https://doi.org/10.1007/s00726-007-0594-0

Yang, M., & Butler, M. (2000). Effects of ammonia on CHO cell growth, erythropoietin production, and glycosylation. *Biotechnology and Bioengineering*, *68*(4), 370–380. https://doi.org/10.1002/(SICI)1097-0290(20000520)68:4<370::AID-BIT2>3.0.CO;2-K

Ye, J., Zhang, X., Young, C., Zhao, X., Hao, Q., Cheng, L., & Jensen, O. N. (2010a). Optimized IMAC-IMAC protocol for phosphopeptide recovery from complex biological samples. *Journal of Proteome Research*, *9*(7), 3561–3573. https://doi.org/10.1021/pr100075x

Ye, J., Zhang, X., Young, C., Zhao, X., Hao, Q., Cheng, L., & Jensen, O. N. (2010b). Optimized IMAC-

IMAC protocol for phosphopeptide recovery from complex biological samples. *Journal of Proteome Research*, *9*(7), 3561–3573. https://doi.org/10.1021/pr100075x

Yeung, Y. G., Nieves, E., Angeletti, R. H., & Stanley, E. R. (2008). Removal of detergents from protein digests for mass spectrometry analysis. *Analytical Biochemistry*, *382*(2), 135–137. https://doi.org/10.1016/j.ab.2008.07.034

Yin, B., Wang, Q., Chung, C.-Y., Ren, X., Bhattacharya, R., Yarema, K. J., & Betenbaugh, M. J. (2018). Butyrated ManNAc analog improves protein expression in Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *115*(6), 1531–1541. https://doi.org/10.1002/bit.26560

Yuan, H. X., & Guan, K. L. (2016). Structural insights of mTOR complex 1. *Cell Research*. https://doi.org/10.1038/cr.2016.10

Yusufi, F. N. K., Lakshmanan, M., Ho, Y. S., Loo, B. L. W., Ariyaratne, P., Yang, Y., … Lee, D. Y. (2017). Mammalian Systems Biotechnology Reveals Global Cellular Adaptations in a Recombinant CHO Cell Line. *Cell Systems*, *4*(5), 530-542.e6. https://doi.org/10.1016/j.cels.2017.04.009

Zhang, F., Sun, X., Yi, X., & Zhang, Y. (2006). Metabolic characteristics of recombinant Chinese hamster ovary cells expressing glutamine synthetase in presence and absence of glutamine. *Cytotechnology*, *51*(1), 21–28. https://doi.org/10.1007/s10616-006-9010-y

Zhang, G., Ueberheide, B. M., Waldemarson, S., Myung, S., Molloy, K., Eriksson, J., … Fenyö, D. (2010). Protein quantitation using mass spectrometry. *Methods in Molecular Biology (Clifton, N.J.)*, *673*(10), 211–222. https://doi.org/10.1007/978-1-60761-842-3_13

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., & Yates, J. R. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chemical Reviews*. https://doi.org/10.1021/cr3003533

Zhou, Y. Y., Li, Y., Jiang, W. Q., & Zhou, L. F. (2015). MAPK/JNK signalling: A potential autophagy regulation pathway. *Bioscience Reports*, *35*(3), 1–10. https://doi.org/10.1042/BSR20140141

Zubarev, R. A. (2013). The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics*, *13*(5), 723–726. https://doi.org/10.1002/pmic.201200451

Zupkovitz, G., Tischler, J., Posch, M., Sadzak, I., Ramsauer, K., Egger, G., … Seiser, C. (2006). Negative and positive regulation of gene expression by mouse histone deacetylase 1. *Molecular and Cellular Biology*, *26*(21), 7913–7928. https://doi.org/10.1128/MCB.01220-06