An Investigation into Approaches and Pathways Relevant to Cellular Flocculation in CHO cells

A dissertation submitted for the degree of Ph.D.

by

Peter Berry, B.Sc.

The work here described was conducted under the supervision of

Dr. Donal O Gorman (Dublin City University)

Dr. Andrew Racher (Lonza Biologics)

Prof. Niall Barron (University College Dublin)

National Institute for Cellular Biotechnology

School of Biotechnology

Dublin City University

Jan 2023

Declaration

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, and 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:

Peter Berry

ID No.: 92165796

Date: 6th Jan 2023

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS USED IN THIS THESIS

ACN	Acetonitrile
ANOVA	Analysis of Variance
APCI	Atmospheric Pressure Chemical Ionisation
ATP	Adenosine Triphosphate
$CaCl_2$	Calcium Chloride
CAM	Cellular Adhesion Molecule
CAR-T	Chimeric Antigen Receptor T Cell
cGMP	current Good Manufacturing Practice
CHAPS	3-cholamidopropyl dimethylammonio 1-propanesulfonate
CHO	Chinese Hamster Ovary
CID	Collision Induced Dissociation
CO_2	Carbon Dioxide
Ct	Critical Threshold
Cy3	Cyanine 3
Cy5	Cyanine 5
Da	Dalton
DAVID	The Database for Annotation, Visualization and Integrated Discovery
DAPI	4’,6-diamidino-2-phenylindole
DE	Differential Expression
DHFR	Dihydrofolate Reductase
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
ECM	Extra Cellular Matrix

EDTA	Ethylenediaminetetraacetic Acid
EI	Electron Impact
ELISA	Enzyme Linked ImmunoSorbant Assay
EMT	Epithelial Mesenchymal Transition
EPO	Erythropoietin
ER	Endoplasmic Reticulum
ESI	Electrospray Ionisation
ETD	Electron Transfer Dissociation
FAB	Fast Atom Bombardment
FAK	Focal Adhesion Kinase
FASP	Filter Aided Sample Preparation
FBS	Feotal Bovine Serum
FC	Fold Change
$FeCl_3$	Ferric Chloride
GFP	Green Fluorescent Protein
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
HAEC	Human Aortic Endothelial Cells
HCP	Host Cell Protein
HEK	Human Embryonic Kidney
HRP	Horse Radish Peroxidase
HSP	Heat Shock Protein
HUVEC	Human Umbilical Vein Endothelial Cells
ICAM-1	Intercellular Adhesion Molecule 1
Ig	Immunoglobulin

IMS	Industrial Methylated Spirits
IVCD	Integral Viable Cell Density
ILK	Integrin Linked Kinase
kDa	Kilodalton
KEGG	Kyoto Encyclopedia of Gene and Genomes
LC-MS	Liquid Chromatography Mass Spectroscopy
LDH	Lactate Dehydrogenase
LED	Light Emitting Diode
MALDI	Matrix Assisted Laser Desorption Ionisation
MAPK	Mitogen-activated Protein Kinase
MBP	Maltose Binding Protein
$MgCl_2$	Magnesium Chloride
$MnCl_2$	Manganese Chloride
MSX	Methionine Sulfoxamine
NAD^+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide (reduced form)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
O/N	Over night
PCR	Polymerase Chain Reaction
PCV	Packed Cell Volume
pDADMAC	Polydiallyldimethylammonium chloride
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PEI	Polyethylenimine
PSD	Particle Size Distribution

PSEA	Protein Set Enrichment Analysis
PSM	Peptide Sequence Matches
PVDF	Polyvinylidene fluoride
Qp	Specific Productivity
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
RT	Room Temperature
RQ	Relative Quantification
SE-HPLC	Size Exclusion High Performance Chromatography
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
STR	Stirred Tank Reactor
TFA	Trifluroacetic Acid
TMB	Tetramethylbenzidine
TSE	Transmissable Spongiform Encephalopathy
UPR	Unfolded Protein Response
VEGF	Vascular Endothelial Growth Factor
WCL	Whole Cell Lysate

ACKNOWLEDGEMENTS

As with all PhDs while it's the candidate who does the work there's a group of individuals operating in the background providing advice and support without whom none of this would have been possible. It's therefore essential that their contribution gets acknowledged.

First and foremost, I would like to thank **Prof. Niall Barron and Dr Andrew (Andy) Racher** for supervising this thesis. This work would not have started without Andy listening to my idea for a flocculating cell line and then becoming an advocate in the essential budgetary meetings that followed or without Niall providing me with the opportunity to be part of his research group. Thanks for saying yes to the proposal and for all your support along the way. You have always been there when needed (even after I started a new role and relocated out of Dublin).

Obviously, I owe a debt of gratitude to the teams in the NICB and School of Biotechnology for all the administrative support that keeps the NICB running and facilitating the MS and HPLC experiments respectively. However, thanks are particularly due to **Dr Padraig Doolan and Prof. Martin Clynes** for facilitating the initial kick off meetings and **Dr Donal O'Gorman,** who as interim director of the NICB never said no to any of my requests. The contribution of **Dr Emma Finlay** for running what I have christened the *R Helpdesk* can't be ignored. Her willingness to accept Zoom invites and troubleshoot/proofread my R code was key to facilitating the analysis presented in chapters 5-7 inclusive. The support of **Dr James Graham, Dr Colin Jacques and Dr Jim Davies** of Lonza Biologics must be acknowledged as must the gift of CHO DXB11 cells by **Prof L Chasin** of Columbia University in New York.

To the CHO gang, **Ricardo**, **Antonio, Berta, Giuseppe, Krishna, Paloma, Prashant, Julie, and Teresa**. Many thanks for our very numerous coffee breaks, laughs, trips, extended chats about data and occasionally intense lunch time debates. I can only say you are friends for life.

Finally, (but most importantly) there's the one person without whom none of this would be possible, my wife **Fiona**. You probably now know more about CHO cells, R code and proteomic analysis than is necessary or you ever wanted to. Your unfailing love and support from day one kept this show on the road. There aren't the words to express my gratitude and the depth of my love for you.

ABSTRACT

Peter Berry - An Investigation into Approaches and Pathways Relevant to Cellular Flocculation in CHO cells

Over the last two decades, viable cell concentrations in industrial processes have reached multiples of 10^7 cells/mL at harvest. Given that many production facilities were built to handle viable cell concentrations in the order of 10^6 cells/mL, cell concentrations of this magnitude have the potential to cause problems during primary recovery activities as existing processes/equipment may not be able to efficiently remove cells from these high cell concentration processes. Flocculation, the ability of cells to clump together is one solution to this issue.

The objective of this thesis was to engineer a flocculating CHO cell line. Three distinct approaches are described. The first approach was the cloning, in CHO cells of four eukaryotic proteins known to promote flocculation. The second approach used inorganic cations as flocculating agents and the third approach utilised proteomics to investigate cellular adhesion in CHO cells.

The nature of the proteins investigated in this thesis lead to challenges in achieving ectopic expression in CHO cells.

While the addition of $FeCl_3$, $MnCl_2$ and $CaCl_2$ did result in improved rates of sedimentation of CHO cells compared to untreated controls without any negative impact on a human antibodies structural integrity, the antibody yield at 62% for $MnCl_2$, 72% $FeCl_3$ and 76% for $CaCl_2$ was not commercially viable.

A differential expression proteomic approach identified a cohort of proteins involved in the *catabolism of amino acids, the catabolism of fatty acids, cholesterol metabolism, protein processing in the endoplasmic reticulum and glycolysis/gluconeogenesis which* were significantly enriched following the adaptation from attached to suspension growth. Additionally, the thesis proposes a potential mechanism which CHO cells developed to combat increased cellular stress during this adaptation process.

This thesis is an describes the first steps towards engineering a flocculating CHO cell line suitable for use with current commercial processes.

AIMS OF THE THESIS

The aims of the work described in the various chapters of this thesis were:

- To design flocculation machinery in CHO cells
- To improve the current understanding of flocculation as primary recovery tool
- Using an approach inspired by nature, to identify and express in CHO cells genes which have the potential to promote cellular flocculation.
- Investigate the use of inorganic cations as flocculating agents for use with CHO cells and their impact on the quality of the cloned recombinant protein.
- Conduct a differential expression (proteomic) investigation of CHO cells following their adaptation to suspension growth in serum free media.

Chapter 3

This chapter examined the potential for five proteins to promote flocculation *in vitro*. These five proteins were:

ABH1 – a class 1 hydrophobin from *Agaricus bisporus*

HFB1 – a class 2 hydrophobin from *Trichoderma reesei*

Flo 5 from *Saccharomyces cerevisiae*

Flo 11 from *Saccharomyces cerevisiae*

Aquaporin 0 - a mammalian protein.

The research aims were as follows:

- Engineer CHO cells to display an inducible flocculating phenotype.
- Express hydrophobin genes in CHO cells as a correctly folded peptide chain.
- Anchor the assembled hydrophobin to the CHO cell surface in such a way that the hydrophobic side was exposed to the aqueous contents of the bioreactor.
- Investigate if there was a difference in the ability of CHO cell lines to express the two classes of hydrophobin.
- Express *S. cerevisiae* genes in CHO cells as correctly folded peptide chains.
- Investigate which *flo* phenotype was better suited to promoting flocculation in CHO cells.

- Subject to the successful expression of these proteins investigate their inducible expression.

Chapter 4

This chapter investigated the ability of calcium chloride, manganese chloride and ferric chloride to flocculate suspension adapted CHO K1 cells at different concentrations and pH and assessed the impact of using these inorganic cations as flocculating agents on the protein quality of a human IgG antibody.

The research aims were as follows:

- Identify concentrations at which inorganic cations be used to flocculate CHO K1 cells
- Assess the impact on antibody yield following flocculation and Protein A/G purification.
- Assess the impact on product quality of a human IgG antibody following flocculation and Protein A/G purification.
- Develop an experimental protocol for using inorganic cations as flocculation agents.

Chapter 5

CHO K1 cells which had been pre-adapted to suspension growth in serum free media were grown as:

- Attached with 5% FBS (A5),
- Suspension with 5% FBS (S5),
- Attached with 0% FBS (A0),
- Suspension with 0% FBS (S0).

Attached and suspension (pre-adapted) cells were compared in a proteomic differential expression experiment as follows:

- A5 v S5,
- A0 v S0.

The research aims were as follows:

- To identify proteins and KEGG pathways whose expression changed between CHO K1 cells grown in attached or suspension culture in the presence of serum (A5vS5).
- To identify proteins and KEGG pathways whose expression changed between CHO K1 cells grown in attached or suspension culture in the absence of serum (A0vS0).
- Identify CHO proteins and KEGG pathways which had potential as future engineering targets for developing a flocculating CHO cell line.

Chapter 6

Chapter 5 used CHO K1 cells which had been preadapted to serum free suspension growth. In chapter 6 attached CHO K1 and CHO DXB11 cells lines which had not been previously adapted to suspension growth were grown as attached cells (A5) and adapted to suspension growth (S5) and compared in separate proteomic differential expression experiments.

- Attached (Serum) v Suspension (Serum) K1 cells
- Attached (Serum) v Suspension (Serum) DXB11 cells

The research aims were as follows:

- Adapt CHO K1 cells from attached to suspension growth.
- Adapt CHO DXB11 cells from attached to suspension growth
- Identify proteins and KEGG pathways in CHO K1 cells whose expression changed following adaptation from attached to suspension growth.
- Identify proteins and KEGG pathways in CHO DXB11 cells whose expression changed following adaptation from attached to suspension growth.
- Identify proteins and KEGG pathways common to both cell lines with the same change in expression following adaptation from attached to suspension growth.
- Identify CHO proteins and KEGG pathways which had potential as future engineering targets for developing a flocculating CHO cell line.

Chapter 7

Suspension adapted cells (chapter 6, from each cell line) were adapted to growth in serum free media and compared in separate proteomic differential expression experiments.

- Suspension (Serum) v Suspension (Serum free) K1 cells
- Suspension (Serum) v Suspension (Serum free) DXB11 cells

The research aims were as follows:

- Adapt CHO K1 cells from growth in media containing serum to growth in serum free media.
- Adapt CHO DXB11 cells from growth in media containing serum to growth in serum free media.
- Identify proteins and KEGG pathways in CHO K1 cells whose expression changed following adaptation from growth in media containing serum to growth in serum free media.
- Identify proteins and KEGG pathways in CHO DXB11 cells whose expression changed following adaptation from growth in media containing serum to growth in serum free media.
- Identify proteins and KEGG pathways common to both cell lines with the same change in expression following adaptation from growth in media containing serum to growth in serum free media.

Chapter 1

Introduction

1.1. Introduction

1.1.1. Chinese Hamster Ovary cells and the production of biologics

Thanks to the discoveries of Jenner (vaccination), von Behring (antibodies), Fleming (the anti-microbial effect of the mould *Penicillium notatum*), society has developed an understanding of the various means by which illness can be prevented/combatted. However, it took until the discovery of DNA, the development of the central dogma (DNA is transcribed into RNA which is translated into protein) and the development of molecular cloning techniques for society to be able to exploit this knowledge and mass produce biological therapies in the quantities necessary to satisfy global demand. The phrase biological therapies or biologics was coined to describe therapeutic proteins e.g., insulin, EPO, antibodies, antibody drug conjugates or other novel therapies e.g., CAR-T which are biologic in nature as opposed to being chemical compounds e.g., aspirin or doxorubicin (a chemotherapeutic drug used in the treatment of breast cancer).

The first recombinant protein licensed for therapeutic use in the early 80's was produced in a microbial expression system. Eukaryotic expression systems based on yeast are also in use but since Genentech produced its recombinant tPA product in Chinese Hamster Ovary (CHO) cells in 1987, the use of expression systems developed for mammalian cell lines to produce biologics has been widely adopted by industry.

Expression of a recombinant protein involves cloning a DNA sequence, which encodes for the desired recombinant product, into a mammalian expression system (Table 1.1). This expression system is then introduced into a cell where it can be integrated into the cells genome. Selective pressure is applied to ensure that only cells which received the expression system can grow. Nutrients are provided to support cell growth and the cells are allowed to grow. The recombinant product is produced by the cell and secreted into the growth media from where it can be purified.

Central to this process is the CHO cell line. First isolated by Puck in the 50's (Puck et al., 1958), and reviewed by Wurm and Wurm (2021) it has become the workhorse of the biologics industry. The reasons for this are several:

Cell Line	Expression Vectors	Selectable Marker	References
NS0	GS Vectors	MSX Cholesterol auxotroph	Bebbington et al. (1992) Branco (2014)
CHO K1	GS Vectors	MSX	www.lonza.com
CHO DUXB11	DHFR intron vector	Both DHFR expression and Hygromycin resistance	Lucas et al. (1996)
DG44	CHEF1	DHFR	Allison (1999)
Serum Free IGF-1 dependent CHO cell line	DHFR bicistronic vector	DHFR	Bianchi and McGrew (2003)
Per.C6	pcDNA3002(Neo)	Neomycin	Jones et al. (2003)
GS-KO (CHO K1 GS knockout)	GS Vectors	Glutamine Synthase	www.lonza.com
Multiple CHO and BHK21	pcDNA3.1-mArg1	Arginase	Capella Roca et al. (2019)
CHOK1SV GS-KO SSI (HD7876)	GS Vector	Glutamine Synthase	Feary et al. (2021)

Table 1.1: **Examples of different mammalian expression systems and selection markers.**

- The fact that CHO cells are genetically unstable (Wurm and Wurm, 2021); this is actually an advantageous trait in industrial settings. This is because a single round of transfection will result in numerous clones with varied productivity rates, growth rates, metabolic profiles and glycosylation patterns all of which aids selection of the optimal cell line for production of the individual biologic.

- It produces glycoproteins that are similar to those produced by human cells. This reduces the chance of proteins produced in CHO cells being folded into the incorrect 3D structure or being rejected by the recipient's immune system. Most biologic therapies are glycoproteins and it's the number and type of attached glycans that play a key role in determining the mode of action of these biological therapies.

- It has been adapted to grow in suspension at higher viable cell concentrations per unit volume than attached cell lines in chemically defined protein and serum free media. This improves the stability and safety profile of the biologic due to the reduced potential for serum contaminants such as extracellular proteases and transmissible spongiform encephalopathies (TSE) such as BSE being introduced into a manufacturing process.

- CHO cells are resistant to human viral infection. This reduces the risk that human viruses could infect and propagate in CHO cells thereby posing a health risk to recipients of any biologic produced in CHO cells.

A recent report in Bioprocess International (Stanton, 2021) estimated that by 2025 the annual global demand (including both commercial and clinical requirements) for these molecules will approach 64,000 Kg, up from 32,000 Kg in 2020. Given that the specific productivity (Qp) of a CHO cell is a measure of how much recombinant protein an individual cell will produce it seems intuitive that one of the ways of increasing recombinant product yield is to increase cell number and cell viability. This is best illustrated with a theoretical example. A culture consisting of $1.8X10^7$ cells/mL at 90% viability will have 16,200,000 cells at harvest in total while a culture consisting of $1X10^7$ cells/mL at 80% viability will have 8,000,000 cells in total at harvest. Therefore, assuming both culture have the identical Qp the culture with $1.8X10^7$ cells/mL will produce approximately twice as much product as a culture consisting of $1X10^7$ cells/mL at 80% viability at harvest as Qp is known to be correlated with culture titre. Indeed, this was

the goal behind the many media design/optimisation programmes. Fed batch stirred tank reactor (STR) processes with a harvest viability in excess of 90% and a viable cell concentration in the region of 10^7 cells/mL are not uncommon at an industrial scale (Dr A. Racher, Lonza Biologics, personal communication). Consequently, increasing three specific variables will govern whether this increase in demand can be met. These are:

1. The specific productivity (Qp) of the host cell line (pg/cell/day);
2. The integral viable cell density (IVCD) of the host cell line; and
3. The overall yield achieved from the manufacturing process (Kg purified product).

The titre of recombinant proteins expressed in CHO cells has increased, at least 1000-fold, from µg/L in the 80's to g/L depending on the process employed. Titres ranging from 10g/L in fed batch (STR) to 27g/L in perfusion systems have been reported for mammalian cell lines (Li et al., 2010 and Kuczewski et al., 2011). This increase in titre and consequently recombinant product yield has been driven by a combination of media design, feed strategy optimisation and cell engineering techniques. However, all these strategies had increasing either Qp or the IVCD of the cells as their core aim. Both Qp and IVCD are correlated with the final product titre and hence yield.

The integral viable cell density can be thought of as the amount of time cells spend producing product and can be broken down into 2 phases. Phase 1 is comprised of how quickly cellular biomass accumulates and phase two of how long the culture remains viable. Qp is a measure of how much recombinant product is produced per individual cell per day (pg/cell/day). Qp values for CHO cells typically range between 20 - 90pg/cell/day (Hacker et al., 2009). The Lewis lab in the University of California at San Diego has calculated the Qp of a single completely senescent CHO clone, as approaching 150pg/cell/day, (Gutierrez et al., 2020). The absolute max Qp for a CHO cell has not been established but as a guide senescent cells like antibody producing B cells can produce 200-300pg/cell/day (Dr N Lewis, personal communication). Bearing in mind that even senescent cells would need some resources for cell maintenance this figure seems unrealistic for a CHO clone producing a recombinant protein.

However, high concentrations of viable cells at harvest can lead to problems during primary recovery when scaled up to the volume required to support market supply. Consequently, the increase in yield which result from increases in product titre can be

lost. The steps involved in a typical biologic production process need to be considered to understand how this loss of biologic product can occur.

1.1.2. A standard manufacturing process for biologics manufactured in CHO Cells

The typical CHO production process is displayed in Figure 1.1. It consists of five stages:

1. Inoculum Expansion: A vial of frozen cells transfected with a gene which encodes for the product (biologic) is thawed and revived. These cells are then initially subcultured into flasks followed by roller bottles and reactors of increasing volume until the production bioreactor is inoculated.
2. Production Bioreactor: The typical commercial CHO manufacturing bioreactor runs for between 13-15 days. During this time various feeds are added to replenish depleted nutrients.
3. Harvest: The contents of the bioreactor are recovered, and the cells/cell debris removed. The resulting liquid, harvest cell culture fluid (HCCF), contains the biologic.
4. Purification: A series of steps to remove process related impurities such as host cell proteins and DNA.
5. Fill Finish: This covers the processes involved in formulating and packaging the product ready for use by the customer.

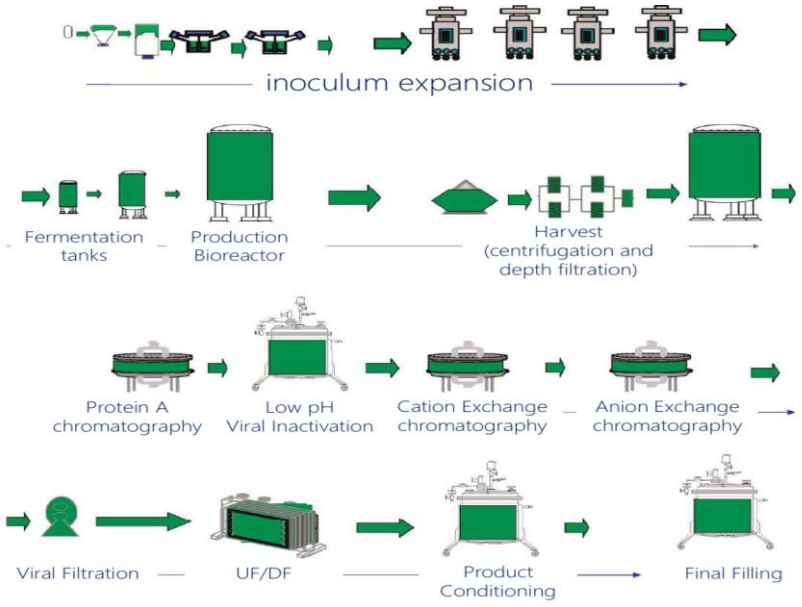
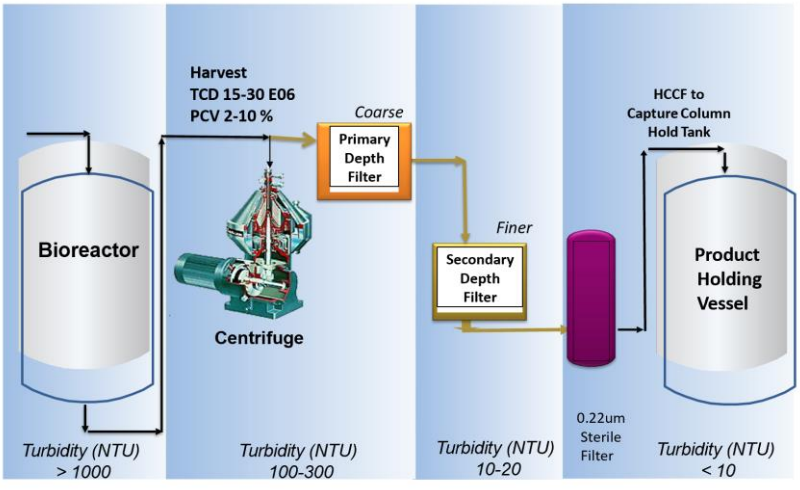


Figure 1.1: The steps involved in a typical fed batch antibody production process (schematic provided courtesy of Prof. M Carrondo, IBET).

1.1.2.1. Primary recovery and packed cell volume

The terms harvest/primary recovery/clarification can be used to cover the steps involving the removal of particulate matter (whole cells, cell debris etc) from the bioreactor contents which has the potential to interfere with purification of the recombinant product. CHO cells and cellular debris are unlikely to have a uniform size. Instead, they will be distributed over a range of sizes. For the clarification process to be capable of removing CHO cells and related cellular debris it needs to work across this size distribution range. As can be seen from Figure 1.1 product purification involves multiple chromatography steps. If particulate matter is not removed prior to the commencement of chromatography steps the chromatographic columns could become blocked resulting in a decrease in final product yield and purity.

A key parameter which needs to be considered during clarification is the packed cell volume (pcv). The term packed cell volume is a measure of the proportion bioreactor volume that comprises of CHO cells expressed as a percentage. The trend in industry for increased cell numbers has led to bioreactors with increased pcv. This parameter is important as pcv can affect the efficiency of disk stack centrifuges and filters in clarifying the bioreactor contents. These are the two techniques most used by industry to clarify the bioreactors contents (Figure 1.2).



Taken from Risk Mitigation Through Innovative Filtration, Jonathan Romero *API Biologics Technical Operations*

Figure 1.2: **The steps involved in primary recovery and the level of clarification (decrease in NTU) achieved at each step.**

1.1.2.2. Disk stack centrifuges

Disk stack centrifuges operate under continuous flow. The contents of the bioreactor are continuously fed from the bottom of the centrifuge into the central bowl which rotates at high speed. As the centrifuge bowl rotates a centrifugal force is exerted on cells and other particles contained in the feed stream. The centrifugal force is selected such that cell and cell debris settle but the recombinant product is retained in the centrate. Cells accumulate as a sludge at the edge of the centrifuge bowl while clarified culture fluid flows up through the centrifuge and can be collected or diverted for further processing. Once the volume of solids inside the bowl reaches a critical volume, they are discharged in a process known as desludging. It follows that increases in the pcv of the bioreactor contents lead to a reduction in the length of time required for the bowl to fill and consequently an increase in the frequency of desludging. There exists a pcv value above which disk stack centrifuges will be constantly desludging and this leads to a loss of recombinant product as the feed stream doesn't remain in the centrifuge long enough for clarification to occur. This pcv value obviously varies based on centrifuge design and capacity.

1.1.2.3. Filtration

Filtration can be defined as the separation of particles from a liquid suspension by use of a porous barrier/filter. Based on their size particles either pass through or are retained by the filter. Particles which are smaller than the diameter of the filter pores pass through, while those whose diameter is larger than the filter pores were retained by the filter. When applied to primary recovery operations the pore size is chosen to retain cells and cellular debris while allowing the recombinant product pass through the filter.

It is not the intention of this introduction to explain or derive the equations governing filtration. It is sufficient to say that as particles are retained by the filter a cake builds up on the filter surface. The thickness of this filter cake creates a resistance to filtration until such time as no further liquid passes through the filter and filtration ends. Pcv could play a role in the rate at which the filter cake forms as increased pcv could contribute to the filter cake forming quicker. This could result in a percentage of the bioreactor contents not being filtered.

1.1.3. Flocculation – A potential solution for the clarification of cell lines with high viable cell concentration at harvest

A potential solution to the primary recovery challenges associated with increased cell densities and high pcv values could be flocculation of the bioreactor contents prior to primary recovery. Flocculation of cells can be defined as the accumulation of the cells in one mass or a propensity for the cells to form clumps. From a bioprocess perspective flocculation involves the cells coming out of suspension and forming flocs or flakes. Flocculation has played a role in several industrial processes, most notably drinking/wastewater treatment and in the brewing industry and can happen either spontaneously (brewing) or due to the addition of a clarifying agent (wastewater treatment plants). CHO cells don't spontaneously flocculate so flocculating agents are typically added to the bioreactor broth (Riske et al., 2007; McNerney et al., 2015).

1.1.4. Mode of Action of Flocculants

The process of cellular flocculation can be divided into the following steps (:

1. Addition and dispersion of the flocculating agent to the bioreactor broth.
2. Adsorption of the flocculant onto the cells.
3. An equilibrium being formed.
4. Floc formation due to collisions between cells which have adsorbed the flocculant onto their surface. These flocs are predominantly formed by one of two mechanisms. These are "electrostatic patch" or "charge neutralisation bridging" (Roush and Lu, 2008). However, Demir et al. (2020) in their review of flocculation mechanisms for microalgae harvesting expanded on these mechanisms to include "compression of the electric double layer", "patch flocculation" and "sweeping." I see no reason why these three additional flocculation mechanisms shouldn't apply to CHO cells. I also see no reason why some of these mechanisms, particularly charge neutralisation and patch or charge neutralisation and bridging shouldn't occur in a synergistic fashion.
5. Floc break up can occur depending on the degree of agitation and consequent shear forces.

1.1.4.1. Compression of the electric double layer (Figure 1.3a)

Cells are predominantly negatively charged and as such will repel each other. Therefore, in order for cells to flocculate these repulsive forces must be overcome. The distribution of charges surrounding any charged surface can be described by the electric double layer model. Repulsion occurs when the electric double layer from two cells overlaps. There are various iterations of the charged double layer model but the three of most interest are the Helmholtz Double Layer, The Gouy-Chapman Double Layer and the Stern modification. The Helmholtz Double layer model is the simplest explanation. It states that ions of the opposite charge to that of the surface cluster at the surface due to attractive forces in a region known as the Helmhotz layer. However, this model is not accurate as it assumes that ions clustering at a charged surface form a rigid layer which does not occur in nature. It's more accurate to consider that the Helmhotz layer as a region (the diffuse double layer) within which ions diffuse away from the charged surface. This modification was described independently by Gouy (1910) and Chapman (1913). Stern modified Gouy and Chapmans model further by introducing the idea of the Stern layer. This assumes that while it's possible that some of the ions in the diffuse double layer are absorbed to the charged surface, the remainder are distributed over a region away from the charged surface. This region is assumed to be the radius of the ion. This layer has a charge potential known as the zeta potential. It therefore stands to reason that if this Stern layer can be compressed thus reducing the zeta potential to the ability of cells to repel each other will be reduced.

This mechanism has been implicated in the flocculation of microalgae using aluminium (Al^{3+}) ions (Cui et al., 2014). The addition of the cation altered the ionic strength of the microalgae suspension thus compressing the Stern layer. This allows attractive van Der Waals' forces to dominate leading to flocculation.

1.1.4.2. Neutralisation of Negative Surface Charges (Figure 1.3b)

While charge neutralisation also minimise the negative surface charge of the cell and allow attractive van Der Waals' to dominate, it differs from compression of the Stern layer in a key aspect. Surface charge neutralisation involves the adsorption of cationic flocculants directly to the surface of a negatively charged cell and it is this that minimises the negative surface charge as opposed to alteration of the ionic strength of the broth.

Aunins and Wang (1989) reported that this was the mechanism by which a polycationic polymer (poly -L_histidine) flocculated CHO cells.

1.1.4.3. Cell Bridging (Figure 1.3c)

Cell bridging involves a form of surface charge neutralisation. This mode of flocculation is characterised by the same polymer molecule being absorbed onto the surface of two or more cells. This results in the formation of increasingly larger flocs. Key to this mechanism of flocculation is the flocculant concentration. Too little and bridging will not occur whereas too much may result in the cells obtaining an overall positive charge with the consequence that flocculation will not occur.

1.1.4.4. Patch Flocculation (Figure 1.3d)

This mechanism describes the adsorption of small positively charged polymers to the cells surface such that regions or “patches” of positive and negative charges are formed. These patches then act as the sites to attract other cells. Flocs are formed by the interaction of positive and negative patches on neighbouring cells. It is this method that Tomic et al., (2015) attributed the flocculation of CHO cells and process impurities such as DNA by pDADMAC to.

1.1.4.5. Sweeping Flocculation (Figure 1.3e)

This describes the trapping of cells within an inorganic precipitate. While Besson and Guiraud (2013) reported that the addition of NaOH was required to flocculate the microalgae *Dunaliella salina* via the precipitation of magnesium ions present in the culture medium, Chen et al., (2017) and Burgstaller et al., (2018) took a different approach with CHO cells. They used a mixture of calcium and phosphate ions and suggested that the underlying mechanism was thought to be related to co-precipitation of calcium phosphate with cell culture debris and impurities, by interaction of the positively charged calcium with the negatively charged particles

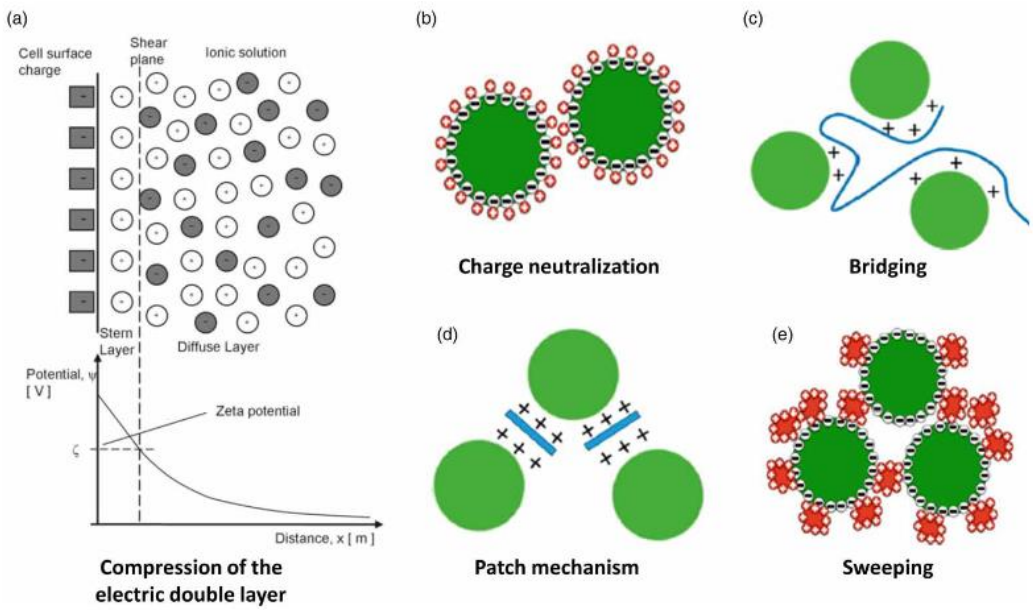


Figure 1.3: **Schematic representation of flocculation mechanisms. A) Compression of the electric double layer, B) Charge neutralisation, C) Bridging, D) Patch Mechanism and E) Sweeping (Reproduced from Demir et al., 2020).**

1.1.5. Chemical flocculants

Charge interactions between phospholipids, sialylated glycolipids and negatively charged glycoproteins present on the surface of CHO cells and a polycationic material is thought to mediate flocculation. This interaction between opposing charges causes the formation of flocs of reduced solubility which further expand in size due to the presence of hydrophobic regions. The overall result is that the flocculated cells rapidly fall out of solution and settle at the bottom of whatever vessel they are in. Consequently, poly-L amino acids (Aunins and Wang, 1989), divalent cations (Romero et al., 2015) and polyacrylamide co polymers, chitosan and polyamines, e.g., PEI, have all found application as flocculants, reviewed by Singh et al. (2016). Additionally, Kang et al. (2013) reported that the partially benzylated poly(allylamine) stimulus responsive polymer developed by Jaber et al. (2011) could be used to flocculate CHO cells. They named this molecule Smart Polymer E or C depending on the polymer's molecular weight.

Two of these flocculants will be briefly discussed in more detail. These are pDADMAC and divalent cations.

1.1.5.1. pDADMAC – a synthetic flocculant

Poly (diallyldimethyl-ammonium chloride) (pDADMAC) is a water-soluble high charge density quaternary ammonium cationic polymer composed of repeating units of diallyldimethyl-ammonium chloride (DADMAC). Its high charge density means it has applications as a flocculant in the wastewater treatment and water purification industries. Merck Millipore have created a cGMP manufacturing process for pDADMAC and have developed an application for it in the flocculation of CHO cells prior to bioreactor clarification by depth filtration. McNerney et al. (2015) reported that the use of pDADMAC as a flocculating agent enabled the development of centrifuge-free primary recovery processes at commercial manufacturing scales while Burgstaller et al. (2018) reported a 4-fold decrease in required filter area to clarify the bioreactor and a 99% reduction in host cell DNA contaminants when using pDADMAC compared to control (no pDADMAC) experiments while maintaining product yield at >97%.

1.1.5.2. Calcium and other cations as flocculants

Flocculating agents aren't confined to synthetic polymers. The role that the divalent cation salt of calcium plays in cell adhesion is well established in the literature, reviewed by Damluji and Riley (1979). Multiple authors have investigated whether calcium chloride on its own (Westoby et al., 2011) or in conjunction with phosphate (Chen et al., 2017 and Burgstaller et al., 2018) is capable of flocculating CHO cells. Magnesium chloride has been used as a flocculant in both wastewater treatment plants (Semerjian and Ayoub, 2003) and the food industry (Li Tay et al., 2006). Additionally, magnesium chloride has been shown to play a key role during cellular adhesion to surfaces (Takeichi and Okada, 1972). Westoby (2011) demonstrated that magnesium chloride caused a reduction in broth turbidity at pH values between pH 5 and pH 6 compared to pH 7 and its use as a chemical flocculant in combination with the commercial flocculant pDADMAC has been investigated (Leng et al., 2015).

The ability of yeast strains which spontaneously flocculate is a key step in the fermentation of beer. In addition to the role the *flo* family of genes play in yeast flocculation there's a wide body of literature describing a role for calcium in the flocculation of brewing yeast (Taylor and Orton, 1973, 1975; Stratford, 1989). Additionally, Stewart and Goring (1976) and Miki et al. (1982), described how magnesium or manganese ions could substitute for calcium in mediating yeast

flocculation. Based on these observations this thesis investigated whether calcium and other divalent cations (iron and manganese) could be used as a flocculating agent for CHO cells. The use of inorganic divalent cations such as these would have an advantage over synthetic polymers as there would be no regulatory requirement to demonstrate clearance by the purification process.

1.1.6. Biological flocculants

1.1.6.1. Cellular adhesion

The ability of cells to adhere to each other and external surfaces is a fundamental biological property. Without cellular adhesion 3D tissue structure would cease to exist and the transmission of signals between cells and their extracellular microenvirnoment would be disrupted. Cellular processes that involve adhesion include cell growth, cell differentiation, embryogenesis, communication, and the immune response.

Several studies have investigated adhesion in CHO cells. It was demonstrated by Brown and Juliano, 1985) that CHO cells bind to fibronectin, a component of the extracellular matrix, via specialised receptors which are members of the integrin family. Cytoskeletal proteins such as talin, vinculin and α-actinin are also known to be involved. More recently Louie et al. (2019) showed that endothelial intracellular cell adhesion molecule 1 contributed to the clumping of cells often observed following the adaptation of CHO cells to growth in serum free media. Klingler et al. (2021) reported a proteomic approach to analysing the surfaceome of CHO cells. Their data demonstrated that the integrins β1 and β3 in addition to CD44, laminin and fibronectin were differentially expressed on the surface of CHO cells prone to aggregation. They further demonstrated that knockdown of integrin β1, using a small interfering RNA knockdown approach, reduced the susceptibility of this population of CHO cells to aggregation. Kumar et al. (2022) linked differential regulation of cell adhesion pathways in CHO cells with increased titre and Qp. They compared two clones of a commercial CHO cell line with two clones of a “in house” CHO cell line expressing the same antibody. Their data showed that the in house cell lines had higher titre and Qp. Furthermore Podoplanin, a mucin type transmembrane glycoprotein, along with other adhesion related proteins involved in focal adhesions (layilin, fibronectin, tensin-1) and podosomes (paxillin and vinculin) were shown to be downregulated in the in house cell line.

1.1.6.2. Cellular adhesion molecules (CAMs) and junctions

Although the cloning and expression of CAMs was not a focus in this project it's useful to briefly discuss them in the context of the significant role they play in cell adhesion. Cellular adhesion molecules (reviewed by Petruzzelli et al. (1999) and McEver and Luscinkas, 2018) fall into 4 classes (illustrated in Figure 1.3a) based on their protein sequence and structure. These are listed in Table 1.2:

However, regardless of what class a CAM belongs to, they all share the same structure. All CAMs possess three domains. These are an intracellular domain, a transmembrane domain, and an extracellular domain. The intracellular domain interacts with the cytoskeleton, the transmembrane domain traverses the cell membrane linking the inside and outside of the cell while it's the extracellular domain which interacts with other CAMs or the ECM (extra cellular matrix) to form a cellular junction depending on which class a particular CAM belongs to.

The interaction of CAMs, either with each other or the cells microenvirnoment, result in the formation of cellular junctions which fall into one of three classes (Figures 1.3b and c)

1.1.6.3. Different CAMs and the type of junction they form

- Communicating junctions, e.g., gap junctions.
- Occluding junctions, e.g., tight junctions; and
- Anchoring junctions. These types of junctions can be further subdivided into junction between cells, e.g., adherens junctions and desmosomes or junctions between cells and the ECM, e.g., focal adhesions and hemidesmosomes. The distinction between adherens junctions and desmosomes or focal adhesions and hemidesmosomes is that focal adhesions and hemidesmosomes involve actin filaments and bind to the ECM whereas adherens junctions and desmosomes do not.

Family	Ligands recognised	Stable cell junctions	Reviewed By
Cadherins	Homophilic interactions	Adherens junctions and desmosomes	Halbleib and Nelson (2006)
Integrins	Extracellular matrix	Focal adhesions and hemidesmosomes	Bachmann et al. (2019)
	Members of the Ig superfamily	No	
Selectins	Carbohydrates	No	Tvaroška et al. (2020)
Immunoglobulin superfamily	Integrins	No	Wong et al. (2012), Shimono et al. (2012), Leshcyns'ka and Sytnyk (2016), Finnegan and Bergstralh (2020).
	Homophilic interactions		

Table 1.2: Cellular Adhesion Molecules (adapted from Cooper, 2000)

Of the three classes of cell junction, it's only anchoring junctions which are involved in cellular adhesion to a surface. Given only cadherins and integrins form stable cell junctions these are the only CAMs discussed in any further detail.

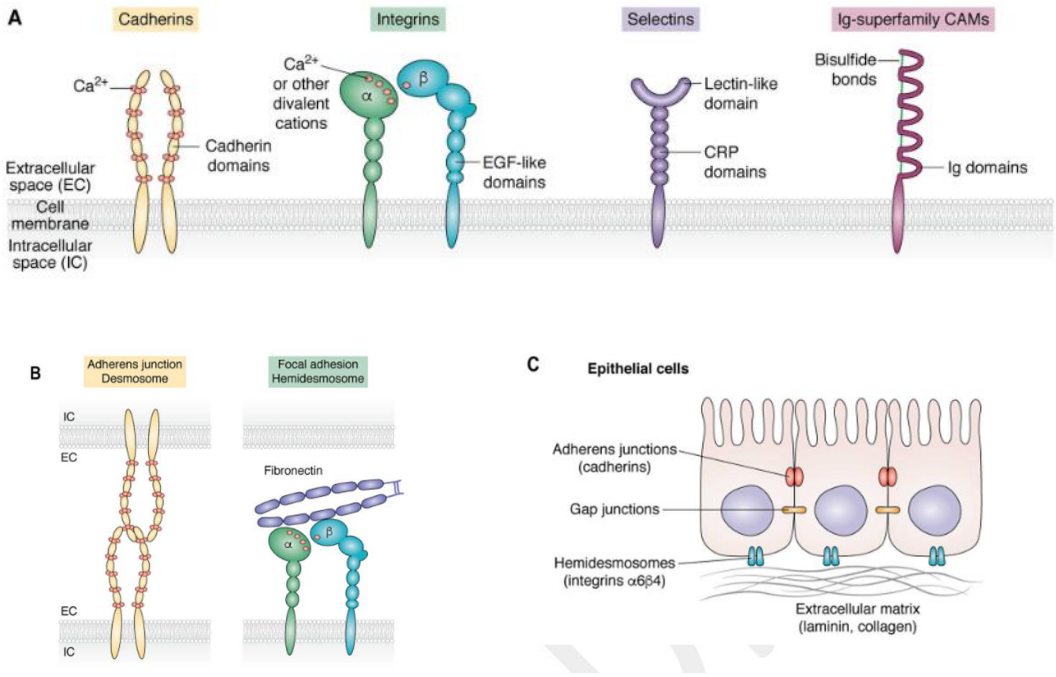


Figure 1.3: **Cellular Adhesion. A) The structure of the four major classes of cell adhesion molecules. B) The different types of cellular adhesions, including adherens junctions, desmosomes, and focal adhesions, hemidesmosomes. C) The different roles cadherins and integrins play in cell adhesion (modified from Janiszewska et al. (2020)).**

1.1.6.4. Cadherins

The cadherin family are calcium dependent CAMs which includes classical cadherins, e.g., E, P and N cadherin, protocadherins, desmogleins, desmocollins and various others which don't belong to a specific category (e.g., CDH proteins). Mature classical and desmosomal cadherins are known to consist of an ectodomain which consists of five extracellular repeats, a transmembrane domain and a cytoplasmic domain which interacts with either β-catenin (classical cadherins) or plakoglobin (desmosomal cadherins). They are known to play a role in cell signalling and the control of cell polarity and proliferation. However, their key function is the formation and maintenance of tissues by reducing the surface tension at the cell-cell interface, initiating rearrangement of the actin cytoskeleton and clustering together to form adherens junctions thus allowing the cell-cell point of contact to increase in size**.**

1.1.6.5. Integrins

Integrins are a protein family composed of two subunits α and β. In vertebrates, such as Chinese hamsters and hence CHO cells, there are 18α and 9β subunits which combine

to form 24 different molecules. As discussed by Bachmann et al. (2019) integrins aren't just involved in cellular adhesion. They are key transmitters of signals in both directions between the cell and its environment and have been shown to play a role regulate gene expression. The "inside-out" signalling model has been proposed to explain the mechanism by which integrins modulate their affinity for their ligands. Conversely the "outside-in" signalling pathway involves proteins such as focal adhesion kinase (FAK), integrin linked kinase (ILK), talin and the formation of focal adhesion complexes. This series of events causes integrins to trigger phosphorylation cascades which impact upon cytoskeletal organisation, establishment of cell polarity, cellular migration, invasion, differentiation, proliferation, and survival.

1.1.6.6. Intercellular Adhesion

Intercellular adhesion can be categorised in two ways:

- By the type of cells which are adhering to each other
  - Homotypic – the attachment of identical cells via adhesion molecules, e.g., the adhesion of two epithelial cells to each other.
  - Heterotypic – the attachment of a cell to a cell of a different type via adhesion molecules, e.g., tumour cells and platelets during cancer metastasis.
- By the CAM which is involved
  - Homophilic – the attachment of an adhesion molecule in one cell to the identical molecule in an adjacent cell, e.g., cadherins in adherens junctions.
  - Heterophilic – the attachment of a plasma membrane adhesion molecule in one cell to a different adhesion molecule in an adjacent cell, e.g., adhesions between integrins present on the surface of a cell and leukocyte intracellular adhesion molecules (a member of the Ig superfamily) present on the surface of endothelial cells.

In the context of engineering CHO cells to flocculate, it would be preferable to focus on CAMs involved in homophilic and homotypic interactions as obviously the cells involved are identical and it would be simpler to engineer and control the expression/interaction of a single CAM rather than multiple different ones. These criteria would make cadherins

the preferred molecule. However, in this project we chose to focus on more novel candidates (Section 1.1.6).

1.1.6.7. Adhesion of Cells to the ECM

Most epithelial cell lines prefer to grow on a surface rather than in suspension. In the context of the cell's native environment this surface is provided by the extracellular matrix (ECM), reviewed by Yue (2014). The ECM is a dynamic environment composed of two types of macromolecules. These are proteoglycans and proteins such as fibronectin, vitronectin, collagen, laminin. The ECM is a source of growth factors and plays a central role in cellular proliferation, migration, polarity, differentiation, apoptosis, and adhesion. In terms of adhesion, it was demonstrated by Pierschbacher and Ruoslahti (1984) that the RGD motif of fibronectin was the motif involved in the binding of fibroblasts to fibronectin. It's now well understood that adhesion of cells to the ECM is mediated by the integrin family of CAMs.

1.1.7. Biological flocculants investigated in this project

Some of the more common industrial applications of flocculation were mentioned in section 1.1.3 but flocculation also occurs in nature. Many strains of microbes (e.g., the bacteria *Staphylococcus aureus* and *Pseudomonas aeruginosa*) and fungal species *Candida, Aspergillus*, and *Cryptococcus*) have evolved to grow as adherent biofilms (a form of flocculation). The advantages of this type of growth includes:

- Aids survival as growing attached to a surface makes it harder to eliminate the microbial population;
- Aids cellular communication;
- Enhances horizontal gene transfer between individual microbes; and
- Provides protection from and resistance to drugs, environmental stresses (e.g., dehydration and ultraviolet light), host immune attacks (e.g., antibodies, complement system, antimicrobial peptides, and phagocytes), and shear forces.

Therefore, molecules exist in nature which could function as flocculants. Examples include the *flo* family of genes in *Saccharomyces cerevisiae* which drive flocculation of yeast cells during brewing. A literature search also suggested the fungal family of hydrophobins and a human protein Aquaporin-0 as candidates for this role.

1.1.7.1. The *Flo* family of genes from *S. cerevisiae*

Given that the term floc derives from the Latin word *floccus* it was reasonable to speculate that the phenomenon of flocculation in brewing was first observed centuries ago. Credit was given to Louis Pasteur for authoring the first report of flocculation in brewing yeast although this has been disputed (Stratford 1992), and in their review (Goossens and Willaert, 2010) credit Stewart and Goring (1976) for the first definition of yeast flocculation. Whatever the origin of the term, what's not in dispute is that the flocculation of yeast plays a critical role in the brewing process (reviewed by Soares, 2011 and Stewart, 2018) and that flocculation is governed by the expression of a specific family of genes (*Flo*) (Lewis et al., 1976). This family of genes encodes a set of cell wall proteins (flocculins) which are capable of recognising and binding to specific carbohydrate structures on neighbouring cells. This process is described by the lectin – carbohydrate model (Miki et al., 1982) and has been demonstrated to be dependent on the presence of calcium cations. Two phenotypes of flocculation in yeast have been described *Flo1* and *New Flo*. These differ in the degree to which different sugars inhibit flocculation. Flo5 (*Flo1* phenotype) and Flo11 (*New Flo* phenotype) were chosen as the candidates for investigation in this thesis. Flo5 has been shown to confer cell-cell adherence and contribute to flocculation (Guo et al., 2000) while Flo11 is responsible for cell-surface adhesions and required for agar invasive growth, pseudohyphae, and biofilm formation (Reynolds and Fink, 2001).

1.1.7.2. Hydrophobins – A unique family of hydrophobic fungal proteins

Hydrophobins are fungal proteins first reported by Wessels et al. (1991a) and were initially discovered in the basidiomycete *S. commune*. To date hydrophobins have been isolated from both ascomycetes and basidiomycetes but don't appear to be present in non-mycelial fungi (for reviews see Wessels (2000) and Wösten (2001)). They are approximately 100 amino acids long, are characterised by the presence of 8 conserved cysteine residues (Schuren and Wessels, 1990) which form 4 disulphide bridges (Wessels, 2000) and their ability to self-assemble into amphipathic membranes at a gas: water interface or on hydrophobic surfaces (Wösten et al., 1993). It's these disulphide bridges which give hydrophobins their solubility in water by preventing spontaneous self-assembly (de Vocht et al., 2000).

Hydrophobins can be spilt into two classes based on their solubility, hydropathy profile and whether they assemble into rodlets or not. Class 1 hydrophobins (e.g., Sc3 from *S. commune*) are soluble in water as a monomer but upon exposure to a gas: water interface they self-assemble into an SDS insoluble amphipathic membrane (Wessels et al., 1991b) and form rodlets (Wösten et al., 1993). This membrane can only be solubilised by using acids such as performic acid (Wessels et al., 1991a) or TFA (de Vries et al., 1993). Conversely, class 2 hydrophobins are soluble in water and 60 per cent ethanol even after self-assembly (Talbot et al., 1996) and don't form rodlets (Wösten, 2001).

The research hypothesis was that CHO K1 cells which expressed hydrophobic proteins on their cell surface would flocculate to reduce the exposure of these hydrophobic surfaces to the aqueous contents of the bioreactor. ABH1, a class 1 hydrophobin from *Agaricus bisporus* and HFB1, a class two hydrophobin from *Trichoderma reesei* were chosen to investigate this hypothesis.

1.1.7.3. Other eukaryotic proteins as flocculants

Given the lack of literature reporting the cloning of yeast and fungal genes in CHO cells it was decided to choose a mammalian protein as a third option. Many molecules have been reported as mediating cellular adhesion, reviewed by Zhong et al. (2015) but it was decided to choose an ocular protein, Aquaporin-0 (reviewed by Roche and Törnroth-Horsefield, 2017). Aquaporins are a family of membrane bound proteins. Aquaporin-0 (AQP0) is the predominant protein in the lens of the eye where it accounts for over 50% of the total membrane protein. Its primary function is in the formation of channels which allow water and other solutes such as glycerol and urea to be transported across the cell membrane. Mutations in AQP-0 has been shown to be responsible for cataracts in mice (Shiels et al*.,* 2001*)*. However, a paper by Kumari and Varadaraj (2009) demonstrated a role for AQP-0 in mediating cellular adhesion and on this basis, it was decided to investigate its potential as a novel biological flocculant in this project.

1.2. Flocculation and primary recovery

Flocculation of the bioreactor contents prior to processing in a disk stack centrifuge can improve centrifuge performance by altering the settling velocity (sedimentation rate) of the bioreactor contents.

The settling velocity (sedimentation rate) of a particle is affected by:

a) The amount of time the particle spends in the centrifuge (residence time);

b) The centrifugal (g) force exerted on the particle by the rotation of the centrifuge (centrifuge speed); and

c) Particle radius. Stokes law states that at a constant speed (i.e., constant g force) an increase in the particle radius would result in faster sedimentation of the particles and hence improved separation efficiency.

This gives us three options to increase the sedimentation rate and hence separation efficiency.

1. The residence time can be increased by reducing the rate at which feed flows through the centrifuge while keeping the centrifugal force constant. This obviously has the effect of increasing the time necessary to process the entire contents of the bioreactor.

2. The centrifugal force can be increased by increasing the speed at which the centrifuge rotates. However, for a given disk stack centrifuge there exists a speed at which particles are not retained in the centrifuge bowl and are ejected with the therapeutic protein in the centrate. This is called breakthrough and should be avoided.

3. The particle size (radius) can be altered.

It is the last point that flocculation of the bioreactor broth aims to affect. By shifting the particle size distribution towards larger sized particles (see Figure 1.4) flocculation has the effect of increasing particle radius and therefore, according to Stokes Law, should result in a primary recovery process with an improved separation efficiency. Additionally, by narrowing the size distribution range and decreasing the pcv value flocculation should decrease the desludging frequency of disk stack centrifuges.

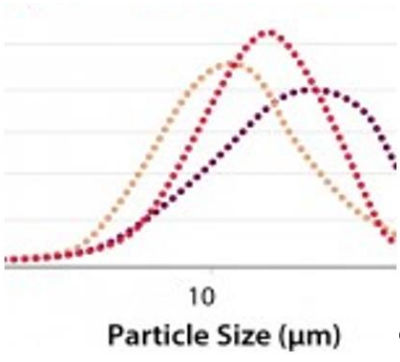


Figure 1.4: **Particle-size distribution (PSD) of a representative CHO cell culture feed stream. Data are shown for unadjusted (pH 7.0) (yellow curve), acid treated (pH 4.8) (red curve), and pDADMAC (0.0375%) (purple curve) treated cell culture feed streams (taken from Singh and Peck, 2014).**

Similarly, flocculation also has the potential to increase the amount of bioreactor liquor which could be filtered per filter unit area by shifting the particle distribution of the liquor towards larger particles thus reducing cake resistance. This was because Endo and Alonso in 2001 demonstrated that in a feed stream comprised of non-spherical particles with a wide size distribution, such as the contents of a bioreactor, the impact on filter cake resistance is significant. One way of visualising this was to imagine a filter cake created from particles of uniform shape and size (perfect spheres). As these particles are retained by the filter and the cake begins to form spaces are left between the particles which fluid can continue to flow through. As the range of particle shape and size increases the smaller particles will start to fill these spaces thereby restricting fluid flow (Dr Jim Davies, Lonza Biologics personal communication). Therefore, it follows that in a filter cake of depth X which is composed of particles with a large size distribution cake resistance will be greater than in a cake of identical depth where the particles are of uniform size or have a reduced size distribution.

The third reason for considering flocculation is that currently there is no clarification process that fits all cell line development and manufacturing scales (**Table 1.3**). This means that the impurity profile of the clarified bioreactor liquor differs at scale. This has the potential to create challenges when designing purification processes which must remain consistent across all scales. Flocculation has the potential to change this situation.

Process Scale	Bioreactor Volume	Clarification Technology
Cell Line Construction	Ambr 15mL or 250mL bioreactors	Benchtop centrifugation or filtration (neutral)
Process Development	up to 10L bioreactors	Bench top depth filtration (adsorbative, e.g., Merck Millistack) Could use a neutral filter but that wouldn't match the adsorbative properties of depth filtration. Consequently, a different range or number of impurities would remain in the filtered material post filtration compared to depth filtration.
Tech Transfer	50-2,000L bioreactors	Two stage depth filtration (adsorbative).
Commercial Manufacturing	Up to 20,000L bioreactors	Centrifugation and one stage depth filtration (adsorbative).

Table 1.3: **Summary of clarification technologies employed at different manufacturing scales.**

The use of bench top versus disk stack centrifuges and charged versus uncharged filters means that the impurity profile of the clarified bioreactor liquor differs by scale. Ideally purification processes would remain constant across all manufacturing scales, but this is not always the case and therefore processes tend to be geared towards the worst-case scenario. This has cost implications and the potential to cause problems when designing purification processes. Since the purification processes used in commercial plants are validated, it's difficult and expensive to make changes to them as the manufacturing scale increases. Accordingly, processes are developed with the worst-case scenario (high impurity load) in mind.

A potential fourth benefit of flocculation is that there is potential for process impurities such as host cell DNA and proteins to get trapped in the flocs thus reducing the burden on the downstream chromatography-based processing steps.

There are, however, disadvantages and potential issues with the use of charged flocculating agents, such as pDADMAC, in biologics production processes. The first of these is that regulators such as the Food and Drug Administration, European Medicines Agency and the Medicines and Healthcare Products Regulatory Agency (in the UK) require evidence that the purification process can remove all traces of the flocculating agent from the final purified recombinant protein. This may involve developing and validating clearance assays specific to the flocculating agent which is an expensive and time-consuming task. A second potential disadvantage to the use of flocculating agents is that they might flocculate the recombinant protein itself thus reducing final yield. A potential solution to both disadvantages was to investigate the possibility of cloning a protein into an expression vector which could be expressed on the CHO cell surface, and which would promote cellular flocculation. As this flocculating agent would be part of the host cell proteome existing platform HCP reduction steps maybe sufficient to clear it. Additionally, as the flocculating agent would be bound to the cell surface as opposed to being present in solution hopefully it wouldn't bind to the recombinant protein to the same extent that charged chemical flocculants might. It is this idea, that a protein which would promote cellular flocculation could be cloned in CHO cells, that this project set out to investigate.

1.3. Growing cells in culture

For mammalian cells to be grown in culture, several requirements need to be met. These include:

- An incubator capable of controlling temperature, pH, osmolality, and humidity within defined parameters;
- An aseptic environment so that contamination is avoided; and
- A growth media capable of providing all the nutrients required by the cell.

Basal media can be defined as media that contains all the nutrients necessary to support the growth of cells. The history of basal media development has been reviewed by

Jayme et al. (1997), Yao and Asayama (2017) and Li et al. (2021). Early media formulations comprised of simple isosmotic salt solutions (e.g., Ringer's, Earle's, and Hank's solutions). Over time formulations that tried to replicate the cells native environment were developed. These formulations could support low density cell growth but required supplementation with additional protein and foetal bovine serum which was used as an undefined source of growth factors and other metabolites.

1.3.1. Adaption to serum free suspension growth

This thesis used CHO K1 and CHO DXB11 cell lines as models to investigate the effects, at a proteomic level, of cellular adaption to serum free suspension growth using a time course label free differential protein expression approach.

Moore and Ulrich (1965) reviewed the growth of mammalian cells in suspension. They distinguished between two methods of suspension culture. The first was where viable but non-proliferating cell suspensions were used to produce viral agents (Frenkel and Waveren, 1938). The second concerned the serial propagation of cells which were actively dividing in suspension culture. They reported that over fifty cell lines had been successfully grown in suspension culture but stated that their attempts to grow fibroblast cell lines in suspension had been unsuccessful. The second method is the method used in biopharma processes today. Murata et al. (1988) and Boraston et al. (1992) reported the successful adaptation of CHO K1 cells to suspension growth by a serial dilution method and manipulating the growth media respectively. In 1996 Sinacore et al. (1996) reported the generation of DUXB11 (DXB11) cells which were preadapted for suspension growth thus avoiding the need to adapt cells to suspension growth each time a vial of cells was revived.

Despite Hamilton and Ham (1977) reporting that CHO cells could be successfully grown in protein free media, the first manufacturing processes at a commercial scale utilised serum in the media and cells were grown attached to the surface of roller bottles. Additionally, animal serum, a poorly defined medium which suffers from high batch to batch variability, was traditionally sourced from foetal cows (FBS).

Using serum to supplement basal media has both benefits and disadvantages.

The advantages of serum include:

- It mimics the cells native environment;
- It's applicable to a wide range of cell types in contrast to modern chemically defined media which, are optimised for specific cell lineages, e.g., Thermo Fisher Scientific sell different media for supporting suspension growth of CHO K1 (CD-CHO or CD-FortiCHO™), CHO DG44 (CD-DG44 Medium) and CHO lineages which are not CHO K1 or CHO S (CD-OptiCHO™). These are different cell lines which are derived from a common progenitor but which have different nutritional needs;
- It's a source of essential hormones, vitamins, cytokines, and transport factors such as transferrin;
- It supports the attachment of cells to a substrate by providing the necessary factors;
- It enhances the ability of basal media formulations to act as a buffer against pH fluctuations; and
- It exerts a protective effect on hybridoma cells (Ramirez and Mutharasan, 1992)

However, the inclusion of serum in media has fallen out of favour with both the biologics industry and regulatory authorities due to the following factors

- It’s expensive and there’s a finite amount available;
- It’s an undefined fluid that suffered from substantial batch to batch variability. This leads to irreproducible cell culture performance;
- Bulk serum proteins such as albumin can interfere with downstream processing steps and negatively impact final product yield;
- There was (and still is) concern about its potential to be a source of mycoplasma, TSE, and other adventitious contaminating agents. However, improved serum collection methods and sourcing from disease free herds can overcome this; and
- Its use goes against the cGMP requirement to fully characterise the manufacturing process and all its constituents.

Accordingly, cell culture media became increasingly more defined until the complex chemically defined, animal component, protein free media used in modern manufacturing processes, were developed (reviewed by Froud, 1999). This review was

followed in 2000 by the publication of a strategy for the adaption of mammalian cells to serum free media (Sinacore et al., 2000).

1.4. Mass Spectrometry and Proteomics

The identification and analysis of individual proteins from complex mixtures, e.g., cells, tissues and biofluids has been revolutionised by the development of highly sensitive mass spectrometers coupled to HPLC systems (LC-MS) such as the Orbitrap from Thermo Fisher or Bruker's time of flight instruments. These instruments when combined with label free quantification methods coupled with intelligent spectral deconvolution tools (Mascot (Perkins et al., 1999) or Sequest (Eng et al., 1994) etc) have allowed the identification and quantitation of proteins whose expression level differs between two or more samples (e.g., control vs treatment). Additionally, modern bioinformatic tools such as Protein Set Enrichment Analysis, (PSEA) (Ulgen et al., 2019) and ontology databases (The Database for Annotation Visualisation and Annotated Discovery (DAVID) (Dennis et al., 2003), Gene Ontology (GO) (Ashburner et al., (2000), and Kyoto Encyclopaedia of Genes and Genomes, KEGG), (Kanehisa et al., 2014 and Kanehisa and Goto, 2000) have allowed for the functional annotation of these proteins. Functional annotation of differentially expressed proteins along with the identification of significantly disrupted biological paths allows a biological context to be assigned to the list of differentially expressed proteins which are the output of any LC-MS/MS proteomic experiment. In this way difference between experimental conditions can be linked to biological impact.

The term proteome is a combination of protein and genome and was coined by a PhD student (Marc Wilkins at Macquarie University) in 1995, Wasinger et al., (1995). Therefore, the cellular proteome describes the complete set of proteins that is encoded in a cells genetic code. Consequently, proteomics is an interdisciplinary field which involves the study of proteins from a protein composition, structure, and activity perspective but which has benefitted from our increasingly comprehensive knowledge of the human and CHO cell genomes. By comparing experimental groups (taken from CHO cells grown under different conditions or displaying different phenotypes) and identifying the proteins which differ between them we may be able to design genetic

interventions to help develop CHO cells with a flocculating phenotype thus making progress towards the central aim of this thesis.

At its most basic level proteomics, as used in this thesis, involves harvesting and lysing cells from the different experimental groups, isolating the protein content of those cells, digesting those proteins into peptides with a protease, separating these peptides on a reversed phase chromatography column, ionising/detecting them in a mass spectrophotometer and applying a range of statistical and bioinformatic tools such as ANOVA, protein databases, gene ontology and enriched pathway analysis to identify changes in the proteome between the two experimental conditions being investigated.

The specific proteomic format used in this thesis was "whole cell lysate bottom up" proteomics. This approach involved digesting the cellular proteins from CHO cell lysates with a protease into peptides. It's these peptides that were separated and identified during LC-MS/MS analysis.

1.4.1. Sample Digestion

A wide range of proteases have been used to break large proteins into smaller peptides for MS analysis. These can be subdivided into nonspecific proteases, e.g., chymotrypsin which cleaves proteins at either aromatic or aliphatic amino acids or highly specific proteases such as trypsin and lys-C which cleave proteins at specific amino acids. Trypsin, which was used in this thesis, cleaves at the C-terminus of arginine and lysine, while lys-C only leaves at the C terminus of lysine (Giansanti et al., 2016). Cleavage by trypsin results in peptides ranging from 700-1500Da. Acidification of the peptides generated by tryptic digestion is used to confer a positive charge to the peptides. One disadvantage of trypsin is that the presence of proline residues or phosphorylation in the vicinity of the cleavage site is known to reduce the digestion efficiency. The use of an excess of trypsin or the combination of multiple proteases (e.g., trypsin and lys-C) is a common strategy for ensuring complete tryptic digestion of the protein occurs.

1.4.2. Mass Spectrometry

It has been estimated that there are in the order of $1X10^{10}$ proteins/single HeLa cell (Milo, 2013). Mass spectrometry is a powerful analytical technique for identifying compounds (both known and unknown) or to elucidate the sequence of molecules e.g., proteins. It achieves this by ionising the sample and measuring its mass to charge (m/z)

ratio. The first instrument capable of separation based on the m/z ratio was developed in 1899. The principle which allowed this separation was based on the fact that ions with different masses could be separated out according to their size as they are passed through a perfect vacuum. Larger masses took longer to traverse the flight tube compared with smaller masses which covered the distance quicker. It's now understood that this discrepancy in flight time was actually attributable to the variation in kinetic energy rather than the mass of the ions. This technology has been modified and improved since then to arrive at the instruments in use today. The instrument used in this thesis was a Thermo Fisher Orbitrap Fusion Tribrid mass spectrometer which has the following properties:

- It combines the best of quadrupole, orbitrap and linear ion trap analysis in a single instrument.
- Precursor selection using a quadrupole mass filter allows the ion trap and Orbitrap mass analysers to operate in parallel for improved sensitivity and selectivity.
- Multiple fragmentation techniques are possible such as collision induced dissociation (CID), higher energy C-trap dissociation (HCD), electron transfer dissociation (ETD) and electron transfer/higher energy collision dissociation (EThCD).
- The subsequent mass analysis can occur in either the low-resolution ion trap or high resolution orbitrap.
- Next generation ion sources and ion optics have increased sensitivity and speed of MS and MS/MS acquisition.
- The ability to carry out ultrahigh resolution between two peaks (up to 500,000 FWHM at 200 m/z) enables molecular weight determination of intact proteins.

It's these characteristics that make the Thermo Fisher Orbitrap Fusion Tribrid mass spectrometer the ideal instrument for performing DE proteomic experiments.

1.4.3. What is a mass spectrophotometer?

At its simplest a mass spectrophotometer (MS) comprises of the following three components:

- Ionisation source
- Mass analyser
- Detector

These components are commonly linked to an HPLC to automate separation of complex peptide mixtures and automate the introduction of samples into the MS (LC-MS). In this case the only variable analysed is the mass to charge ratio (m/z) of the precursor ion which is generated in the ionisation source. LC-MS can be modified by the addition of a second mass analyser to provide LC-MS/MS. This allows for selection and fragmentation of specific precursor ions based on their m/z ratio and in the case of peptides allows the sequence of the selected peptide (precursor ion) to be determined. MS/MS analysis can also result in increased sensitivity due to the increased signal/noise ratio in the second mass analyser.

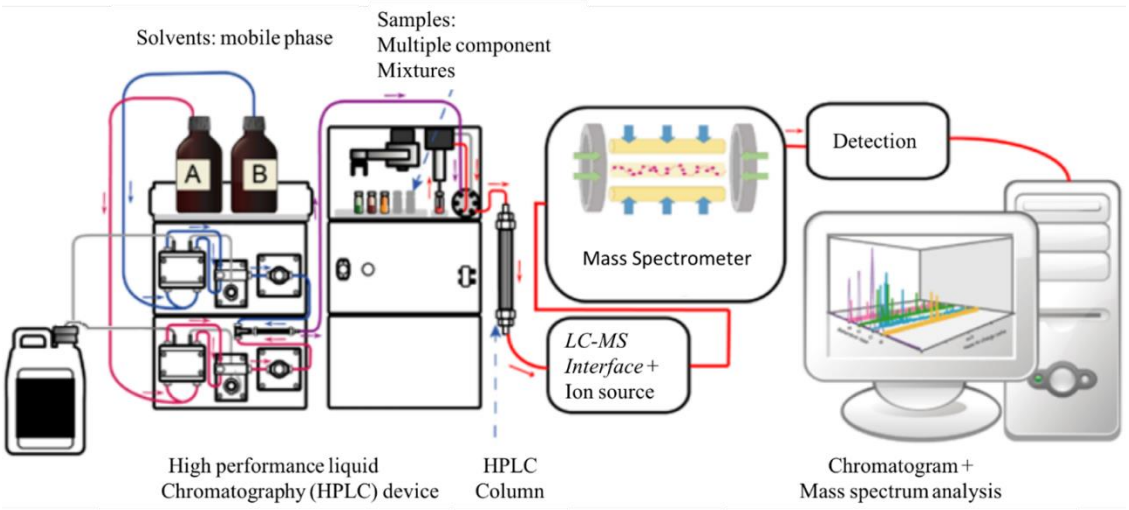


Figure 1.5: **Example LC-MS Layout (https://www.technologynetworks.com/analysis/articles/lc-ms-what-is-lc-ms-lc-ms-analysis-and-lc-msms-348238).**

1.4.3.1. Sample Ionisation

Proteomic based mass spectrometry requires the sample to be ionised in the gas phase. This involves the removal of an electron from the peptide resulting in the formation of a molecular or parent ion. There are many methods of ionisation e.g., matrix assisted laser desorption ionisation (MALDI), electron impact (EI), fast atom bombardment (FAB) and atmospheric pressure chemical ionisation (APCI) but the one used in this thesis was a variant of electrospray ionisation (ESI) called nano electrospray ionisation as it maximizes the effective analyte concentration to give enhanced signal response for dilute samples.

ESI as an ionisation method results in ions with multiple charges. In simple terms this means the ion of interest has picked up an additional proton, which results in an altered mass to charge ratio compared to the native peptide. The maximum charge state of a peptide can be influenced by multiple factors including peptide fragment length and the amino acid composition of the peptide. Amino acid composition is probably more important as histidine (H), lysine (K) and arginine (R) can all be protonated. Peptides which are highly abundant in these amino acids have a greater probability of being multiply charged than peptides which are less abundant in these amino acids. Trypsin tends to produce short peptides most often observed at +2 and +3 charge state although higher charges are possible albeit with a reduced frequency.

1.4.3.2. Fragmentation and MS/MS Analysis

Fragmentation can occur in two locations,

- In source fragmentation; and
- In the collision cell located between the first and second mass analysers.

It's the result of energetically unstable precursor ions breaking down into positive ions and an uncharged free radical. It's this fragmentation pattern that allows the peptide sequence to be elucidated.

Until recently in source fragmentation was considered undesirable, Lu et al. (2017). However, recent publications by Domingo-Almenara et al. (2019) and Xue et al. (2020) have demonstrated that in source fragmentation during ESI in conjunction with enhanced in-source fragmentation/annotation resulted in fragmentation patterns which were consistent with those generated following fragmentation in the collision cell

at a collision energy of 20eV. They validated this approach using a range of metabolites including the amino acids aspartate, glutamine, glutamate, phenylalanine, tyrosine, and tryptophan.

1.4.4. Quantification Approaches Routinely Employed in the Analysis of MS/MS Data

There are two competing approaches to peptide quantification in an MS experiment (Patel et al., 2009). These are labelled, reviewed by Beynon and Pratt (2005), or label-free, reviewed by Pham et al., (2012). Labelled approaches such as stable isotope labelled amino acids in cell culture (SILAC), isobaric tag for relative and absolute quantitation (iTRAQ) and tandem mass tag (TMT) approaches rely on metabolic (SILAC) or chemical (iTRAQ and TMT) labelling. This approach enables the absolute quantification of peptide abundance in a sample. In contrast label-free approaches don't enable the absolute quantification of peptide abundance in a sample. Label-free approaches enable the identification of proteins whose expression profile has changed between different experimental conditions e.g., control, treatment 1, treatment 2 etc by directly comparing the peak intensity of a given peptide ion across all samples. In this way the label-free approach allows the degree of change (the fold change) of each peptide between different experimental conditions (differential expression (DE) analysis) to be determined and the protein expression profile across experimental conditions to be established.

A differential expression label-free approach was chosen in this thesis since label-free analysis doesn't suffer from the complex sample preparation steps, the requirement for expensive reagents and the risk of incomplete labelling which have been associated with label-based approaches. Neilson et al. (2011) reported on study of different methods of protein quantification involving fifty-two participants and mixtures of known concentration. This study demonstrated that while stable isotope labelling gave greater consistency between replicates compared to label-free methods, results obtained by label-free methods were closer to the expected values compared to methods employing stable isotope labelling. Neilson et al. (2011) also stated that "it's generally agreed that label free quantitation can estimate protein abundance more accurately than gel-based methods and can observe the differential abundance of proteins over a larger dynamic

range than labelling techniques." Despite this the label free approach is not without its limitations. It's well established that multiple injections of the same sample will display differences in peak intensities for a given peptide ion between injections. This variation in peak intensity needs to be accounted for, otherwise errors in quantitation will occur. While these differences are often attributable to technical variation during sample preparation, they can be minimised by making multiple injections of the same sample vial (technical replicates) or employing a normalisation strategy. Additionally, any drift in peak retention time or m/z ratio will lead to misidentification of peptides. The effect of this is minimised by performing sample alignment prior to comparing peaks. Once again, any inaccuracies at this stage will lead to increased variability between samples thus negatively impacting on peptide identification and quantitation of the degree of change between experimental conditions. The use of software packages, for label free analysis, which employ some of these strategies is discussed in Section 1.4.5.2.

1.4.5. Peptide and Protein Identification

1.4.5.1. The m/z Ratio

The mass to charge (m/z) ratio of an individual ion is the basis on which peptides generated by tryptic digestion can be identified by mass spectrometry (MS1). Each peak present in an ion chromatogram represents a peptide whose m/z value can be established. This peptide will consist of multiple amino acids and therefore has a mass which is equivalent to the sum of the mass of each of the individual amino acids which make up the peptides sequence. The use of a tandem mass spectrometer (MS2) allows the unambiguous identification of a precursors peptide sequence. MS/MS analysis uses collision induced dissociation to fragment a precursors peptide ion (MS1) of defined m/z value. Optimisation of the kinetic energy within the collision cell results in the cleavage of a single peptide bond thus producing two fragments. The cleavage site, however, varies stochastically such that the output of a tandem mass spectrometer (MS2) is a series of peaks (Figure 1.6). These peaks correspond to all N-terminal fragments superimposed over all C-terminal fragments. Analysis of the difference in m/z ratio between neighbouring peaks allows identification of the amino acids which make up the selected peptides sequence.

1.4.5.2. Databases for Peptide and Protein Identification

While calculating the difference in m/z ratio for the different peaks generated by MS/MS analysis could be done by hand for a single protein, this may not be feasible when analysing complex mixtures such as cell lysates. To this end, databases, containing all tryptic peptides coded by a species genome, exist to aid peptide and protein identification *in silico*. Search engine algorithms such as Sequest or MASCOT can be used to interrogate a given database and based on certain user defined parameters (database name, precursor and fragment mass tolerance, the protease(s) used and an allowance for missed cleavage sites) report all possible peptides based on the specified protease. Comparison of the observed profile, from a MS/MS spectrum, with the theoretical profiles contained in a database allows the elucidation of a peptides sequence. Similarity scores are calculated between the theoretical and observed profile for each peptide. This similarity score in conjunction with a minimum algorithm scoring threshold provides confidence in the identification. Finally, once all peptides have been identified this information is used to infer which proteins were present in the sample being analysed.

While protein databases from non-CHO species (e.g., Human) have been successfully used to assign protein identifications to CHO proteins (Meleady et al., 2012a) the reporting of a draft CHO-K1 genome by Xu et al., (2011) allowed the release of an annotated NCBI reference transcriptome based on this draft CHO-K1 genome in 2012. In their 2012 paper, Meleady et al. (2012b) demonstrated that the use of a CHO derived protein databases resulted in an increase of approximately 50% in the number of protein identifications when compared to the use of protein databases based on other species. Protein identification databases derived from the CHO genome have continued to evolve as evidenced by the release of a further NCBI reference transcriptome in 2014 which was based on a Chinese hamster species draft genome (Lewis et al., 2013). The proteomic studies described in this thesis utilised a proteogenomic draft annotation of the Chinese hamster *(Cricetulus griseus)* to match mass spectra and peptide information. This proteogenomic draft annotation of the Chinese hamster genome was experimentally annotated using RNA-Seq, proteomics, and Ribo-Seq data (Li et al., 2018).

The use of these databases in conjunction with commercial software packages such as Progenesis QI for Proteomics or Maxquant LFQ has simplified the analysis of label free experiments. Progenesis QI for Proteomics was the software used in chapters 5, 6 and 7 and was supplied by a division of Waters (www.nonlinear.com). It displays each individual sample in a graphical format with the m/z ratio on the x-axis and retention time on the y axis. It has the advantage that alignment and normalisation of the samples happens automatically once the raw MS data has been imported into the software. An experimental design based on a one way ANOVA with a p-value ≤ 0.05 is defined by the user. This enables the comparison of peak intensity for each peptide between different experimental conditions and results in the identification of peptides whose expression level differs significantly between experimental conditions. Peptides identified in this manner are then picked, quantitated and exported for protein identification using Proteome Discoverer and the appropriate database. Following this, peptides which have been identified as contributing to a protein's identification are imported back into Progenesis QI for Proteomics and the individual abundance of each peptide contributing to a protein's identification are summed. In this way differences in each protein abundance, between experimental conditions, can be determined and expressed as a fold change across experimental groups.

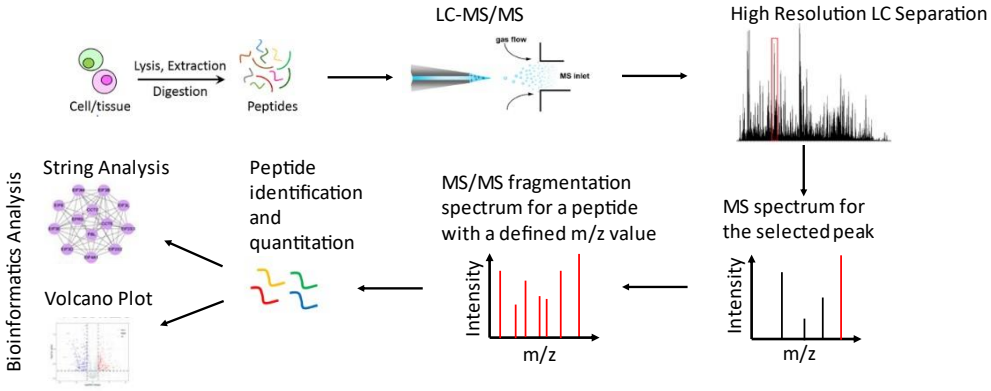


Figure 1.6: Schematic showing the process by which proteins are extracted from cells, digested into peptides, and separated by LC. Following this separation, the MS peptide spectrum is obtained. Individual (precursor) peptides are further fragmented resulting in the acquisition of a MS/MS fragmentation spectrum for each peptide ion.

Chapter 2

Material and Methods

2.1. General Cell Culture Techniques

2.1.1. Ultra-Pure Water

Ultrapure water was prepared using a Milli-RO 10 Plus system (Millipore) and it was used for the preparation of all media and buffer solutions if not otherwise specified. The process involved the pre-treatment of water with using a proprietary method followed by purification by reverse osmosis. Water was sterilised prior use for 20 minutes at 121°C and 1 bar with saturated steam if used for cell culture purposes.

2.1.2. Cell Culture Cabinets

All procedures (cell culture and reagent preparation) which required aseptic conditions were performed in a class II laminar air flow cabinet (Holten). Before any use the cabinet was turned on and the air flow allowed to equilibrate for 15 min to ensure a sterile, aerosol free air flow. Immediately prior to use both the operator's gloves and the laminar cabinet were sprayed with a solution of 70% Industrial Methylated Spirits (70% IMS) and wiped down. All items (e.g., cell culture flasks, pipettes, tip boxes, media bottles, reagents etc were sprayed with 70% IMS before being placed inside the laminar cabinet. Post use all items were removed, the cabinet was wiped with 70% IMS and allowed to run for 15 minutes before being turned off.

Monthly cleaning of laminar cabinets and cell culture incubators involved disinfecting all surfaces with a chlorine-based detergent (Virkon®) for 15 minutes followed by washing with water and disinfecting with 70% IMS.

All cell culture cabinets were certified on an annual basis to ensure that they were operating in line with the manufacturer's specifications.

2.1.3. Cell Culture Incubators

Adherent cells were maintained at 37°C in an atmosphere of 5% CO_2, 80% relative humidity in static incubators. Suspension cultures were maintained at these conditions in an ISF1-X (Climo Shaker) Kuhner incubator shaking at 170 rpm for routine growth and 130 rpm during suspension adaptation experiments. Incubators were cleaned on the same schedule and in the same way as for cell culture cabinets.

2.1.4. Sterilisation of Cell Culture Reagents and Equipment

All thermal stable solutions and pipette tips were sterilised for 20 minutes at 121°C and 1 bar with saturated steam prior to usage. Thermal labile solutions (e.g., DMSO, FBS, Trypsin/EDTA) were filtered through 0.22µm filters (Millipore, Millex-GV, SLGV-025BS) with the help of sterile syringes or purchased as sterile solutions.

2.2. Subculture of Cell Lines

To avoid cross contaminations only one cell line at a time was handled inside the laminar hood. Prior to using a different cell line, a 15 minute gap was established, and all surfaces were resterilised with IMS. All cell lines used for experiments were cultured without the use of antibiotics. To maintain cells in mid exponential growth phase they were passaged into fresh media every 3-4 days. The cell lines used in this thesis are tabulated in Table 2.1

2.2.1. Maintenance and Propagation of Adherent Cell Lines

Adherent cells were routinely cultivated in T75 flasks (Corning®) in a Steri-cycle® CO_2 incubator with a HEPA-filter (Thermo Scientific). Cells were grown at 37°C, 5% CO_2 in a humidified atmosphere. For sub-cultivation media was removed into a sterile waste container using a pipette. The cell monolayer was washed with 10mL of sterile PBS to remove residual media and serum. PBS was discarded and 5mL prewarmed trypsin/EDTA solution was added to the flask. The cell/trypsin suspension was incubated for five minutes or until cells had detached at 37°C in an incubator. Cells were visually monitored using a microscope to confirm detachment. 5mL of serum containing media was added to neutralise the trypsin/EDTA solution and cells were transferred into sterile 30mL universal containers and pelleted for five minutes at 300Xg in a Thermo Electron Jouan C3i multifunction centrifuge. The supernatant containing trypsin/EDTA solution was decanted off and the cell pellet was resuspended into 1 mL of fresh, prewarmed media containing the appropriate concentration of serum. Cells were counted using the Guava (Section 2.3.2). The required number of cells was then added to prewarmed media (final volume 15 mL) and this mixture was transferred into a new T75 cell culture flask. Typically, flasks were either seeded at $2.5X10^5$ cells/flask and split every three days or $1X10^5$ cells/flask and spilt every fourth day.

Cell Line ID	Source	Culture Format	Culture Media	Media Supplements	Chapter Used In
CHO S	Purchased from ThermoFisher	Preadapted to suspension	Forti CHO (routine and transfections)	None	Chapter 3
CHO K1	In House	Preadapted to suspension	BalanCD CHO Growth A (routine) CHO-S-SFM-II (Transfections)	None	Chapters 3, 4 and 5
CHO DXB11	Gift From Prof. L. Chasin, Columbia University	Adherent (adapted to suspension in this project)	aMEM (Adherent) BalanCD CHO Growth A (Suspension)	100x HT and FBS	Chapters 6 and 7
CHO K1	Purchased from ECACC	Adherent (adapted to suspension in this project)	aMEM (Adherent) BalanCD CHO Growth A (Suspension)	FBS	Chapters 6 and 7

Table 2.1: Cell Lines, Culture Conditions and Media using during this project.

2.2.1.1. Preparation of 0.25% Trypsin, 0.5mM EDTA Solution

For the sub cultivation of adherent cell lines, a Trypsin/EDTA (TV) solution was used. The Trypsin/EDTA solution consists of 0.25 % (w/v) Trypsin and 0.02 % (v/v) EDTA. The solution was prepared by dilution of 2.5% (w/v) Trypsin (Gibco®, catalogue No. 15090046) solution and a sterile 0.5M EDTA solution with sterile phosphate buffered saline (PBS) to a final volume of 500mL. This resulted in a 0.25% Trypsin, 0.5mM EDTA solution in PBS. The TV solution was aseptically aliquoted and stored at -20°C.

2.2.2. Maintenance and Propagation of Suspension Cell Lines

Cells were cultivated in a Kühner Climo-shaker incubator (ISF1-X) with a process environment set to 37°C, 170 rpm **(130 rpm for suspension adaption experiments)** and 80% relative humidity. For routine growth TPP 50mL TubeSpin® (Helena BioSciences) bioreactor tubes for 5mL batch cultures or 250mL Erlenmeyer shake flasks (Corning®) were used. For routine sub cultivation cells were counted using the Guava (Section 2.3.2). The required number of cells was calculated, and cells were seeded at an initial cell density of $2X10^5$cells/mL in 50ml TubeSpin® bioreactor tubes (5mL culture volume) or 250 mL Erlenmeyer shake flasks (30mL culture volume). Typically, suspension cells were subcultured every three days before they exceeded $2X10^6$ cells/mL. This was to ensure cells remained in an exponential growth phase.

2.2.2.1. Cryopreservation of Cell Lines for Long Term Storage

Long term storage of mammalian cell lines was at -196°C in the liquid phase of liquid nitrogen. Attached cells were trypsinised, resuspended in serum containing media and spun down at 170Xg for 5 minutes. The cell pellet was resuspended in 1mL of basal media (e.g., aMEM for attached DXB11 cells and attached CHO K1, Balan CD for suspension adapted CHO K1 cells and Forti CHO for CHO S cells) without DMSO or serum and counted using the Guava (Section 2.3.2). The volume was adjusted such that a final cell number in the range of 4-$6X10^6$ cells/mL was obtained. An equal volume of cell suspension and prechilled 2x freezing media (e.g., aMEM supplemented with 20% DMSO and 20% FBS for attached cells or media supplemented with 20% DMSO for suspension cells) were mixed to reach lx final concentration and cells were transferred into cryopreservation vials (CRYO.STM, Greiner Bio One, 201151). This resulted in a final cell concentration of between 2-$3X10^6$ cells/mL per cryopreservation vial. DMSO prevents the formation of ice crystals. Cells were transferred immediately into a -20°C freezer and stored for 2 to 3 hours before they were stored in the liquid phase of liquid nitrogen at -196°C.

2.2.3. Cell Line Revival

Given the cytotoxic effect of DMSO in high concentrations long incubation times in media containing DMSO was avoided. Cells were thawed by pipetting up and down in pre-warmed media and immediately transferred into appropriate media for suspension culture or cultivation in monolayer. Suspension cells were transferred

into 250mL Erlenmeyer flasks (30mL) while T75 cell culture flasks were used for adherent cells with a culture volume of 15mL. To remove dead cells and debris the culture media was exchanged 24 hours after thawing and viability was assessed.

2.3. Cell Counting and Viability Determination

Three methods were used for counting cells and assessing viability.

2.3.1. Use of a Haemocytometer (Trypan Blue Exclusion method)

Prior to cell counting an aliquot of cell suspension was taken either during subculture of adherent cells or by direct sampling of suspension cultures. Briefly an aliquot of cells was diluted in PBS, and then mixed with an equal volume of Trypan Blue dye. 10µL of the cell/trypan blue suspension was applied to a hemacytometer and the number of stained (blue) and unstained cells in the corner square was counted (marked in blue in Figure 2.1). The number of stained and unstained cells in each of the four corners were counted and divided by 4 to obtain the average. Application of the appropriate correction factor (10,000 or 10^4) for the hemacytometer and correction for dilution enabled the number of viable cells/mL and % viability to be obtained. Samples were assayed in duplicate and the average of the two counts calculated.

Note:

The haemocytometer in use at the NICB had a depth of 0.2mm rather than 0.1mm

Viability (%) = [1 - (No of stained cells / No of unstained cells)] * 100

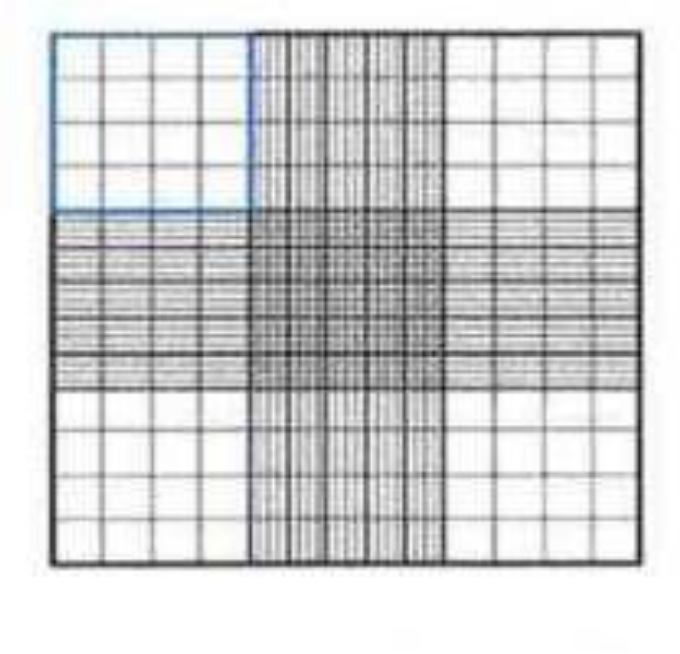


Figure 2.1: **Standard Haemocytometer. Taken from https://www.abcam.com/protocols/counting-cells-using-a-haemocytometer.**

2.3.2. Guava Viacount® Assay

The GUAVA EasyCyte® bench-top flow-cytometer (Merck) was used for high throughput 96 well plate-based cell viability assays. Briefly an aliquot of cells was diluted in PBS, and then mixed with an equal volume of ViaCount solution (Section 2.3.2.1) before being read on a Guava instrument. The method used two DNA binding dyes with different membrane permeability properties and absorption spectra to distinguish between viable and non-viable cells. Viable CHO cells adsorbed the LDS-751 dye but excluded Propidium Iodide (PI) dye as PI cannot permeate the CHO cell membrane. Non-viable cells are known for having compromised membrane integrity and hence can take up both dyes.

2.3.2.1. Preparation of ViaCount solution

400 µL of Propidium Iodide (1.0 mg/mL in H_20, Thermo Fisher Scientific, P3566)

200 µl of LDS-751 (2.5mg/mL in DMSO, Thermo Fisher Scientific, L7595)

100 mL of PBS

Stored at 4 °C protected from light.

2.3.3. Use of the Ceedex XS Analyser

The Cedex XS Analyser was a semi-automated image-based cell counting system which used an optical slide rather than a hemacytometer to count cells and was based on the well-established Trypan Blue exclusion method for determining cell viability. Samples were prepared by mixing 20µL of cell suspension with an equal volume of Trypan Blue dye. 10µL of this mixture was loaded into a slide and the slide loaded into the Cedex XS analyser. Each slide could analyse up to 8 samples and samples were analysed in duplicate. Cell counting and graphical analysis of the results were performed automatically by the Cedex XS software. The advantage of this system was that it also provided an estimation of cell size, and this was useful when assessing the degree of cellular flocculation.

2.3.4. Routine testing of cell lines for contaimination by *Mycoplasma species*

All cells were routinely assessed for *Mycoplasma* contamination using an indirect staining method to detect *Mycoplasma* contamination. *Mycoplasma* testing was performed at the NICB by Ms Justine Meiller. In context of this project no *Mycoplasma* contamination was detected in any cell line used.

The method described here is taken from the relevant NICB SOP and involved passaging cells for a minimum of three passages in antibiotic free, drug free, selective agent free media. 5mL of conditioned cell culture media was removed from 2/3 days old cultures and the cells pelleted by centrifuging in a Thermo Electron Jouan C3i multifunction centrifuge for 5 minutes and excess media was decanted off to a separate container. This was stored at -20°C until required. Normal rat kidney (NRK) fibroblasts are used as an indicator cell line. *Mycoplasma* negative NRK cells are seeded on coverslips and allowed to attach overnight. Following confirmation of attachment of the NRK cells, 1mL of conditioned cell culture media was applied to duplicate coverslips and allowed to grow in a 5% CO_2 humified incubator @ 37°C until ~50% confluent.

Once the NRK cells had reached ~50% confluency the coverslip was washed twice with PBS, followed by one wash with a 50:50 solution of cold PBS/Carnoys fixative (1 part glacial acetic acid to 3 parts methanol). Washed NRK cells were fixed to the coverslips by the application of 2mL of Carnoys fixative to each coverslip for 10 min after which the fixative was decanted off and coverslips allowed to dry overnight .

Dried coverslips were stained with 2mL of Hoechst 33258 stain (50ng/mL, Sigma

Catalogue No. B2883) for 10 min, washed three times with sterile distilled water and dried off with a medi wipe. From this point on the work proceeded without direct light to limit quenching of the fluorescent stain. The coverslips were mounted onto slides with a drop of mounting medium (50% (v/v) glycerol, 0.02M citric acid, 0.06M disodium phosphate). The slides were examined first under phase contrast to check cell integrity and then under fluorescence to check for the presence of *Mycoplasma* species. Test samples were compared to positive and negative slides to aid the analysis of slides. The Hoechst 33258 dye stained the nucleus of the NRK cells and any extra-nuclear DNA (such as *Mycoplasma)* which may be present. *Mycoplasma* infected samples had small fluorescent bodies present in the cytoplasm of the NRK cells. Occasionally these small fluorescent bodies were visible outside the NRK cells.

2.4. Molecular Biology Techniques

2.4.1. Gene Synthesis and Plasmid Reconstitution

The DNA sequence of ABH1, HBF1, Flo5 and AQP0 were obtained from GenBank. These were synthesised, ligated into the plasmid pcDNA3.1 and supplied as 4µg of lyophilised plasmid. Plasmids were reconstituted by centrifuging the vials at 6000Xg from 1 minute at 4°C followed by the addition of 20µL of sterile nuclease free water (Ambion®, AM9932) and vortexed for 1 minute. The concentration of the reconstituted plasmids was determined as per Section 2.4.2, diluted to 100ng/mL in nuclease free water, aliquotted and stored at -20°C. Long term storage of plasmid stocks at -80°C involved transforming competent *E. coli* cells with the plasmid (Section 2.4.2), isolating single colonies, and preparing 10% glycerol stocks

The DNA sequence encoding for the AQP0-de2GFP fusion protein was created in house using the plasmid SW-BB. This plasmid encodes for de2GFP which lacks an ATG start codon and is based on pcDNA3.1. The gene for AQP0 was modified by adding Kpnl and Sgsl restriction site using a high fidelity *Taq* polymerase and the 5' and 3' ends respectively. The modified AQP0 gene was then ligated into the SW-BB vector on a Kpn1/Sgs1 fragment such that ATG of the GOI was in frame with the sequence encoding for de2GFP. This provided assurance that if transfected CHO cells fluoresced when assayed in the Guava Express Plus programme the GOI was being translated.

The plasmids used in this project are tabulated in **Table 2.2.**

Cell Line ID	Plasmid	Source	Gene	Chapter Used In
CHOS	pcDNA3.1	GenScript	ABH1, HFB1, AQP0	Chapter 3
CHO K1	pcDNA3.1	Genscript	ABH1, HFB1, AQP0, Flo5	Chapter 3
CHO K1	Modified SW-BB	Ricardo Valdes Bango Curel	AQP0-GFP Fusion	Chapter 3

Table 2.2: **Plasmids and Cell Lines used during this project.**

2.4.2. Nucleic Acid Quantification

For the quantification of DNA and RNA concentrations the NanoDrop,™1000 spectrophotometer was used. Prior to use the NanoDrop,™1000 was initialised using 2µL H_20 and it was blanked with 2µL of sample diluent. 2µL of the sample was dropped directly on the optical measurement surface of the spectrophotometer and the absorbance at a wavelength of 280nm obtained. From this the concentration of the DNA or RNA sample could be calculated. Furthermore, information about the purity of a DNA or RNA sample is provided by absorbance ratios at 260/280nm and 230/260nm. Pure RNA has a 260/280nm ratio of ~2.0 and pure DNA has a 260/280nm ratio of ~1.8. Absorbance at 230nm is viewed as an indicator of other contaminants. Pure/good quality nucleic acids were expected to have an A260/A280 between 1.8-2.0 and an A260/A230 between 1.8-2.3. For accurate measurement, each sample was measured 2-3 times and the average was calculated. When possible, samples were always measured in the appropriate range, 20ng/µL - 3000ng/µL, for accuracy.

2.4.3. Bacterial Transformation

For plasmid amplification, bacterial transformations were performed in a 1.7mL micro centrifuge tube using 50µL of chemically competent *E. coli* One Shot Top 10 cells (Invitrogen, C404010) and 2ng plasmid DNA. The cells/plasmid mixture was incubated on ice for 30 minutes prior to being subjected to a heat shock at 42°C for

40 seconds in a water bath and immediately returned to ice for 2 minutes. 250µL of S.O.C. medium (Invitrogen) was added to the transformed cells which were incubated for 1 hour at 37˚C in a Grant-Bio ES-20 Environmental Shaker-Incubator shaking incubator at 220 rpm. Cells were then pelleted by centrifugation in an Eppendorf 5424 microfuge at 3,000Xg for 5 minutes. 200µL of supernatant was discarded and cell pellets re-suspended in the remaining medium. The cells were plated on pre-warmed LB-Agar, containing 100µg/mL ampicillin, by spreading 20µL of bacterial suspension over the plate. Plates were incubated overnight at 37°C, lid down. Colonies could be observed after the incubation, and plates could be stored at 4°C for up to 2 weeks. After overnight incubation single colonies were used to inoculate 5mL of LB broth containing 100µg/mL ampicillin as a selection marker. These were incubated overnight at 37°C at 225 rpm in a Grant-bio-ES-20 Environmental Shaker-Incubator and were used as the source material for plasmid purification (Section 2.4.3).

Note:

The genotype of *E. coli* One Shot® TOP 10 cells was:

F- mcrA A(mrr-hsdRMS-mcrBC), (D80lacZAM15, AlacX74, recAl, araD139 A(araleu)7697, galU, galK, rpsL, (StrR), endAl, nupG.

2.4.4. Plasmid DNA Extraction (Mini and Maxi Preps)

Plasmid DNA was isolated prior to transfection of CHO cells using Mini or Maxi Prep Kits.

For mini preps (Thermo Scientific™ GeneJET, catalogue number K0503) single colonies picked from agar plates were inoculated in 5mL LB media containing selection agent. For maxi preps (Invitrogen™, catalogue number K210007) the culture volume was 50mL. Both methods are based on the same principle with the exception that plasmids prepped using maxi kits are endotoxin free and can be used for transfections. Briefly, cells are harvested and resuspended in RNase containing buffer. The cells are lysed using alkaline lysis followed by the separation of protein and cell debris from plasmid DNA using centrifugation. The plasmid DNA is bound to positively charged silica membranes and washed to remove endotoxins, lipids, and polysaccharides before elution. The plasmid DNA was isolated following the manufacturer's protocol prior to being quantified as per Section 2.4.2.

2.4.5. High-Fidelity Polymerase Chain Reaction (PCR)

For high-fidelity amplification of specific DNA fragments from plasmid DNA template for further cloning, the Phusion® High-Fidelity PCR Master Mix (Thermo Fisher, F360S) was used. All reactions were set up in ice. Primer concentrations ranged from 0.2-1μM. Template concentration ranged between 20-200ng of DNA. The protocol followed for this type of reaction was:

10.0μL	Phusion Master Mix (2X)
10.0μL	Forward Primer (10μM)
1.0μL	Reverse Primer (10μM)
XμL	Nuclease Free H_2O
XμL	Template DNA
20.0μL	Total volume

PCR cycling conditions varied according to the template/fragment to be amplified and were adjusted starting from the following and taking into account the manufacturer recommendations (see **Table 2.3).**

Step	Temperature	Time	Cycles
Initial Denaturation	98°C	30 min	1
Denaturation	98°C	5-10 s	30
Annealing	User determined	10-30 s	
Extension	72°C	15-30 s/kb	
Final Extension	72°C	15-30 s/kb	1
Hold	4°C	Hold	

Table 2.3: PCR conditions used during this project.

2.4.6. Transient Transfection of CHO Cell Lines

All transfections were double vector transfections. Vector 1 was pcDNA3.1 containing the gene of interest. Vector 2 was either N44 which contains a destabilised isoform of GFP (de2GFP) or pcDNA3.1 containing the gene encoding for eGFP. The vectors were mixed in the ratio of 9:1 (i.e., 450ng vector 1 and 50ng vector

2) prior to dilution and mixing with transfection reagent.

2.4.6.1. TranslT®-X2 Mirus for plasmid DNA

Parental suspension adapted CHO K1 cells were seeded at $5X10^5$ cells/mL 24 hours prior to transfection in 250mL flasks containing Gibco CHO-S-SFM II media (Gibco®). The complex formation was performed in CHO-S-SFM II media without any supplements in 1.7mL microcentrifuge (Eppendorf) tubes. Cells were transfected with a total of 500ng of DNA/$1x10^6$ cells. Firstly, the DNA was adjusted to a final volume of 350µL of media and 1µL Mirus transfection reagent (Mirus Bio® LCC, MIR 6003) was added for every µg of plasmid DNA used. To allow the complex formation the reagent mix was incubated for twenty minutes at room temperature. During this time the cells were centrifuged, washed with fresh media, counted, and plated in twenty-four well plates (Corning®) at $1X10^6$ cells/well. The complexes were added drop wise to the cells and incubated overnight. After twenty-four hours GFP expression was assessed and 1mL samples were taken for RNA isolation. The remaining 1mL was transferred to TPP 50mL TubeSpin® bioreactor tubes and 4mL of Balan CD media (Irving Scientific) was added. These cultures was used after a further seventy-four hours growth to assess genotypic and phenotypic expression of the cloned gene.

2.4.6.2. Freestyle™ Max Transfection Reagent

Parental CHO S cells were seeded at $5X10^5$ cells/mL twenty-four hours prior to transfection in 250mL flasks containing Forti CHO media (Gibco®, catalogue). The complex formation was performed in OptiPRO™ (Gibco®, catalogue) serum free media without any supplements in 1.7mL microcentrifuge (Eppendorf) tubes. Cells were transfected with a total of 1µg of DNA/$1X10^6$ cells. DNA was adjusted to a final volume of 750µL of media. In a separate 1.7mL microcentrifuge (Eppendorf) tube 1µL of Freestyle™ Max reagent/µg DNA was made up to 750µL with media. For complex formation the diluted Freestyle™ Max reagent was added dropwise to the DNA and mixed by swirling the pipette tip. The Freestyle/DNA mixture was incubated for a maximum of twenty minutes at room temperature. During this time the cells were centrifuged, washed with fresh media, counted, diluted and 5mL added to TPP 50mL TubeSpin® bioreactor tubes at $1X10^6$ cells/mL. The complexes were added drop wise to the cells while swirling the TPP tube and incubated overnight. After twenty-four hours GFP expression was quantified. This was used

after a further seventy-two hours growth to assess genotypic and phenotypic expression of the cloned gene.

2.4.6.3. Modified Transfection Protocol for Transfecting Adherent CHO K1 Cells

Eight chambers of a Borosilicate #1 chambered cover glass system (Labtech, 155411) were seeded with 400µL of parental CHO K1 cells at $6.25X10^4$ cells/mL in Gibco CHO S II SFM supplemented with 10% FBS. Chambers were incubated overnight in a humidified static incubator @ 37°C, 5% CO_2 prior to transfection with plasmid and Mirus TransIT®-X2 transfection reagent.

On the day of transfection, the media and any unattached cells were removed and 350µL of fresh Gibco CHO S II SFM without FBS was added to each chamber. Plasmid encoding for the fusion protein and Mirus reagent were prepared as previously described (Section 2.4.6.1), mixed, complexes allowed to form and 50µL of the plasmid/Mirus mixture was added dropwise to four chambers. Chambers were incubated overnight in a humidified static incubator @ 37°C, 5% CO_2 prior to visualising on a Nikon Ti2-E fully motorised inverted research microscope (Section 2.4.6.4).

The remaining four chambers were transfected with N44 only as anegative control.

2.4.6.4. Detection of Fusion Protein Expression in Transfected CHO K1 Cells

Individual cells were visualised on a Nikon Ti2-E fully motorised inverted research microscope equipped with fast filter wheel and camera triggered piezo stage insert at 60X under oil immersion with the appropriate filters for GFP and DAPI fluorescence. Filters (Semrock) consisted of individual excitation filters in a SpectraX light engine, quad LED filter cubes for DAPI, GFP, Cy3 and Cy5, with individual emission filters in the filter wheel. GFP was excited at (peak) 470nm and emitted at 509 nm. Nuclei were counterstained blue with NucBlue™ Live Ready Probes™ Reagent (Gibco, R37605) which excites and emit at the same wavelengths as DAPI. Fast Z stack images were obtained and overplayed through control with NIS Elements, deconvolved using Richardson-Lucy deconvolution algorithms and combined to form a 3-dimensional volume view image.

2.4.7. Total RNA Extraction from Transfected CHO K1 and CHO S Cells

Total RNA from CHO K1 and CHO S cells was isolated using TRIzoL" Reagent (Thermo Fisher Scientific, 15596026). Cell suspension containing $2.0x10^5$ - $5.0x10^6$ cells was pelleted at 300Xg in a Jouan C3i multifunction centrifuge for five minutes and supernatant was discarded. Cell pellets were lysed by adding 1mL of room temperature TRIzoL™ reagent and mixed by pipetting up and down until homogenous solution was obtained. Samples were incubated for five minutes at room temperature for immediate processing or stored at -80°C for later processing (after thawing them on ice). 200µL of Chloroform (Sigma-Aldrich, 496189) per mL of TRIzoL was added to each sample. Samples were vortexed vigorously for fifteen seconds, then incubated for two to three minutes at room temperature. For phase separation, the TRIzol™/Chloroform mix was centrifuged in a Sigma 1-1EPK bench top centrifuge at 12,000Xg at 4°C for fifteen minutes. Aqueous phase (top) was carefully transferred to a clean micro centrifuge tube while avoiding the transfer of any interphase or organic layer. For RNA precipitation, 500µL of isopropanol (Sigma, 19516) per mL TRIzol™ was added to the samples and incubated for ten minutes at room temperature. Samples were centrifuged in a Sigma 1-1EPK bench top centrifuge at 12,000Xg for ten minutes at 4 °C. After decanting the supernatant, RNA pellets were washed twice using 1mL of 75% Ethanol (prepared with DEPC-treated Nuclease Free water) and centrifuged at 7,000Xg for five minutes at 4°C. After the second wash, all supernatant was carefully removed using a pipette, and pellets were air dried for 5-10 minutes, before being re-suspended in 20µL of pre-warmed (55°C) Nuclease-free water (Ambion®, AM9906). Samples were further incubated at 55°C for 10 minutes, then quantified using Nanodrop® (2.4.2) and stored at -20°C (short term) or -55°C (long term).

2.4.8. Detection and Relative Quantitation of Gene Expression by RT-qPCR (SYBR)

2.4.8.1. Primer design

In order to quantify gene expression, specific primers for the gene(s) of interest were designed using the primer-blast function on the NCBI site (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and ordered from Integrated DNA Technologies (IDT). Upon arrival, lyophilized primers were spun down, re-

suspended in nuclease-free water for a stock final concentration of 100µM following manufacturer's instructions, mixed and spun down. A working stock for each pair of primers was prepared by dilution the original stock 1:50 for a concentration of 2µM. Stocks of individual primers or pairs of primers were prepared. All primer stocks and working stocks were stored at -20 °C. Before use, primers were thawed on ice, vortexed for 10 - 15 seconds and spun down for proper re-suspension.

2.4.8.2. cDNA Synthesis

For gene expression quantification, total RNA had to be reverse transcribed into complementary DNA (cDNA) using the enzyme reverse transcriptase. The Invitrogen High-Capacity cDNA Reverse Transcription Kit (lnvitrogen, 4368814,) using up to 2µg of total RNA in a final reaction volume of 20µL was used. Depending on the reaction volume 10X RT buffer, 25X dNTPs (100nM), 1µL MultiScribe™ reverse transcriptase (50U/mL) and 10X random primers were mixed and added to the total RNA sample. To reach the final reaction volume of 20µL nuclease free H_2O was added as necessary. For the first strand synthesis primers were annealed at 25°C for 10 minutes. The reverse transcription was performed at 37°C for 120 minutes and the reaction was stopped at 85°C for 5 minutes followed by a hold step at 4°C. All reactions were carried out in a G Storm Thermo cycler system. The cDNA was stored at -20°C until required.

2.4.9. Quantitative PCR (qPCR)

For RT-qPCR MicroAmp optical 96-well plates (Applied Biosystems, 4346906) with Fast SYBR® green 2x master mix (Applied Biosystems, Fast SYBR® Green Master Mix, 4385612) were used. The reverse transcribed cDNA was diluted to obtain a concentration of 10ng/µL The master mix was prepared using 10µL 2x Fast® SYBR green master mix, 200nM of each primer targeting the gene of interest and 20ng cDNA. To reach the final reaction volume of 20µL nuclease free H_2O as necessary. Samples were prepared and measured as triplicates. The reaction was carried out with an Applied Biosystems 7500 Fast Real-Time PCR system using 40 cycles. Cycles contained a denaturation step at 95°C for 30 seconds and an annealing/elongation step at 60°C for 15 seconds.

To reduce variability among samples due to reaction preparation, a master mix containing both the Fast SYBR® Green master Mix (2X) and the cDNA template, with a 20% excess to account for volume losses, was prepared for each condition to be

analysed, taking into account the number of primers sets to be used. The mix was then split according to the number of primers, and each primer mix was added. Each reaction was prepared to be ran in triplicate and loaded in a MicroAmp® Fast Optical 96-well Reaction Plate (Applied Biosystems), which was sealed using an optical adhesive film, and centrifuged to ensure all sample was at the bottom of the plate before the measurements were performed.

2.4.9.1. Methods for relative quantification of mRNA after RT-qPCR

In order to determine the relative amounts of transcripts between samples/conditions, Ct values were used to calculate the Relative Quantification (RQ) using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). This method allows the calculation of the relative amounts for each transcript in relation to a reference control sample/condition. The method also uses Ct values of an endogenous control known not to change between samples or conditions. GAPDH was routinely used in this work as an endogenous control. RQ values were calculated using the following formula:

$$RQ = 2^{-\Delta\Delta Ct}$$

Where $-\Delta\Delta Ct =$

$$(Ct_{target} - Ct_{endogenous})_{condition\ 1} - (Ct_{target} - Ct_{endogenous})_{Reference\ condition}$$

2.4.10. Determining Transfection Efficiency for CHO K1 and CHO S Cells

To evaluate transfection efficiencies (% of fluorescent cells) or transgene expression (Mean Fluorescence Intensity - MFI) when using the GFP reporter gene, the *Express Plus* program of the GUAVA EasyCyte® bench top flow cytometer was used. Gates for dark and bright cells were defined using dark untransfected cells and bright positive control cell populations. All samples were analysed using the same settings for comparability of the obtained results. Cells for analysis (Section 2.4.6.1 and 2.4.6.2) were diluted in PBS and plated in U-shaped bottom ninety-six-well plates to a concentration between 10-500 cells/µL where possible. Typically, 3000 cells were analysed per sample.

2.5. Plasmid Cloning/ Maps

In silico maps of the different constructed plasmids were created using the following license-free software:

• SnapGene Viewer - used to generate the plasmids maps

2.6. Biochemistry & Proteomic Methods

2.6.1. Cell Lysis

Cell lysis was performed by adding 500µL of lysis buffer (7M urea, 2M thiourea, 4% CHAPS, 30mM Tris pH8.5 and 1x protease inhibitors) to the cell pellet. Thermo Fisher Scientific Halt™ Protease Inhibitor Cocktail (100X) was the inhibitor mixture used. Cells were pelleted as described in Section 2.6.3.5. Lysis was aided by passing the cell pellet through a blunt end 21-gauge needle 10 times. The samples were incubated on a circulating rotator for 1 hour at RT followed by centrifugation in a Sigma 1-1EPK bench top centrifuge at maximum speed (17,000Xg) and 4°C for 15 minutes. The supernatant was removed to a fresh tube and total protein content quantified by the Bradford method (Section 2.6.2). 100µg of protein was aliquoted into separate tubes and if necessary, the final volume was adjusted to 100µL with 50mM ammonium bicarbonate. The samples were stored at -80°C for long term storage or -20°C overnight .

2.6.2. Protein quantification - Bradford Assays

For the measurement of protein concentration BSA (Thermo Scientific, 23208) pre diluted standards were used. These standards were further diluted 1 in 2 with lysis buffer to give final protein concentrations of 1.0mg/mL, 0.75mg/mL, 0.5mg/mL, 0.25mg/mL, 0.125mg/mL, 0.063mg/mL and 0.031mg/mL. This was necessary to account for the fact that the standards come formulated in PBS whereas cell lysates are in sample lysis buffer. Sample diluent was included in all assays as a blank. A range finding plate was prepared using undiluted samples. 5µL of each standard, sample or blank was transferred in triplicate into a flat bottomed 96-well plate. 250µL Coomassie Brilliant Blue G25O Quick Start™ Bradford Protein Assay (Bio-Rad, 5000201) was added to each well and incubated for ten minutes. The absorbance was read using Multiskan™GO Microplate Spectrometer at a wavelength of 595nm. Based on the results of this plate, if it was necessary, samples were diluted in PBS such that they fell in the middle of the standard curve and the assay repeated.

2.6.3. SDS-PAGE and Western Blotting Analysis

2.6.3.1. Reduced SDS-PAGE

Either 4-12% Bis Tris gradient gels (NuPAGE™Novex™, Thermo Fisher Scientific, NP0322BOX) or 12% Run Blue™ Bis-Tris gels (Expedeon, NBT01212) run under denaturing conditions in an XCell SureLock™Minicell electrophoresis chamber (Thermo Fisher Scientific, EI0001) were used. The samples were prepared by mixing 4x NuPAGE™ LDS sample buffer (Thermo Fisher Scientific, NP0008) and 20x reducing solution (IM DL-DTT solution) (Sigma, 646563) with cell lysate such that the final concentration of sample buffer and reducing agent was 1x. This solution was then denatured and reduced by incubation at 95°C for 5 minutes. Samples were allowed to cool down before being centrifuged for 30 seconds in a microfuge. 10µg of total protein was loaded on the gel for each sample in a maximum volume of 30µL and the gel was run for 30 minutes at 200V (constant voltage) using lx MES running buffer (Expedeon, NXB70500). 10µL of a prestained molecular weight ladder (Thermo Fisher Scientific, SeeBlue™ Plus2 pre-stained protein standard LC5925 or Expedeon, RunBlue™ prestained marker NXA05160) was included to allow size estimations. To maintain denaturing conditions 0.5mL NuPAGE® antioxidant (Thermo Fisher Scientific, NP0005) was added in the upper chamber of the electrophoresis chamber. The gel was stained with Instant Blue protein stain (Expedeon, catalogue ISB1L).

2.6.3.2. Non Reducing SDS PAGE

Non reducing SDS PAGE analysis was performed using the same gels, buffers and molecular weight ladders as reducing SDS PAGE (Section 2.6.3.1). The sample preparation for non-reducing SDS PAGE differed to reducing SDS PAGE and was as described below:

- The samples were prepared by mixing 4x NuPAGE™ LDS sample buffer (Thermo Fisher Scientific, NP0008) with cell lysate such that the final concentration of sample buffer was 1x. This solution was then denatured by incubation at 95°C for 5 minutes. .Samples were allowed to cool down before being centrifuged for 30 seconds in a microfuge. 10µg of total protein was loaded on the gel for each sample in a maximum volume of 30µL and the gel was run for 30 minutes at 200V (constant voltage) using 1x MES running

buffer (Expedeon, NXB70500). 10μL of a prestained molecular weight ladder (Thermo Fisher Scientific, SeeBlue™ Plus2 pre-stained protein standard LC5925 or Expedeon, RunBlue™ prestained marker NXA05160) was included to allow size estimations.

2.6.3.3. Protein Transfer from Polyacrylamide gels to PVDF membranes

The Thermo Scientific™ Pierce™ Power Blotter (catalogue no. 22834) consisting of the Pierce Power Station with blotting software and the Thermo Scientific™ Pierce™ Power Blot Cassette (catalogue no. 22835) for rapid semi-dry transfer of proteins from polyacrylamide gels to PVDF membranes was used. Once electrophoresis had been completed, the gel was equilibrated in a 1X tris-glycine transfer buffer (Bio-Rad, 161- 0734) for approximately ten minutes. Four sheets of Whatman 3mm filter paper (Whatman, 1001824) were soaked in freshly prepared transfer buffer. Two sheets of filter paper were placed on the anode plate of a Thermo Scientific™ Pierce™ Power Blot Cassette (catalogue #22835). Air pockets were removed from between the filter papers by rolling with a clean pipette. PVDF membrane (GE Healthcare, 10600021) was cut to the size of the gel and was activated by soaking in 100% methanol for two seconds and equilibrated using the transfer buffer prepared earlier. This was placed on top of the filter papers on the anode plate. Air pockets were once again removed. The gels were then placed onto the membrane, and two additional sheets of transfer buffer- soaked filter paper were placed on top of the gel. All air pockets were removed, and excess transfer buffer was removed from the anode plate. The blotting cassette was closed by placing the cathode plate on top of the anode plate and the cassette was inserted into the Pierce Power Station. Proteins were transferred from the gel to the membrane using the pre-programmed settings in accordance with the manufactures' recommendations. Transfer efficiency was assessed by visually examining the blot after transfer for the presence of the pre stained molecular markers. Additionally, the gel was stained with Instant Blue protein stain (Expedeon, catalogue #ISB1L). Reduced staining intensity compared to gels which weren't transferred was expected and taken as evidence that protein transfer had occurred.

2.6.3.4. Western Blotting of PVDF membranes

Following protein transfer, the PVDF membrane was blocked for one hour at 4°C on

a rocking platform using 1X Odyssey PBS blocking buffer (Licor, 927-900001). The membrane was probed with primary antibody (Mouse anti 6X His tag, Thermo Fisher Scientific, MAI-21315, Rabbit anti AQP 0, Abcam, ab134695) prepared in blocking buffer with 0.05% Tween 20 at recommended dilutions, and incubated overnight at 4°C on a rocking platform. The membrane was then washed 4 times with PBS containing 0.1% Tween 20 for a total of 20 minutes. Secondary antibody, prepared in in blocking buffer with 0.05% Tween 20 1/6000 dilution of either anti-mouse (H+L) (Licor, 926-32212D) or anti-rabbit (H+L) (Licor, 925-68073D), was added and the membranes incubated for 1 hour at RT. Post incubation with secondary antibody, the membranes were washed four times thoroughly in PBS containing 0.1% Tween 20 for a total of 20 minutes followed by a final wash for 5 minutes with PBS (no tween). Blots were scanned using the Odyssey Infrared Imaging System.

The Odyssey instrument was turned on 10 minutes prior to use and allowed to warm up. The scanner surface was wetted with Milli Q H_2O and blots were placed face down on the scanner. Scan resolution and quality were set as medium. The XY co-ordinates describing the region to be scanned were selected and the fluorophore channel setting adjusted according to the fluorophore present on the secondary antibody:

- Anti-mouse (H+L) (Licor, 926-32212D) emitted in the green channel @795nm and the channel settings were 700nm (0.5), 800nm (5).
- Anti-rabbit (H+L) (Licor, 925-68073D) emitted in the red channel @694nm and the channel settings were 700nm (5), 800nm (0.5).

Molecular weight markers prestained with bromophenol blue could be visualised in the red channel due to the overlap of the bromophenol blue emission spectra and this wavelength. Images were exported as .tiff files.

Given the purpose of these experiments was to show presence or absence of protein expression rather than differences in protein expression levels an internal control (eg GAPDH) was not included on these blots. However, the appropiate positive controls as described in Section 3.3.3 and Section 3.6 respectively were included.

2.6.3.5. Sample Preparation for MS Analysis

Samples were prepared by Filter-Aided Sample Preparation (FASP) as described by

Coleman et al. (2017). 100 µg of total protein from whole cell lysates (WCL) was used as the starting material. Each experimental parameter was analysed as four replicates, i.e., four cell culture flasks were setup and processed separately. All reagents were prepared in MS grade H_2O unless otherwise stated.

2.6.3.6. Whole Cell Lysis

Adherent Cells

Adherent cells were washed with 10mL sterile PBS. The PBS was decanted off and the cell monolayer removed by scrapping with a sterile cell scrapper. Cells were counted and at least $1X10^6$ cells were pelleted at 300Xg in a Thermo Electron Jouan C3i multifunction centrifuge for 5 minutes and the supernatant was decanted off. The cell pellet was washed twice with cold PBS and the cell pellet was either stored at -80°C until required or lysed immediately.

Suspension Cells

Cells were counted and at least $1X10^6$ cells were pelleted at 300Xg in a Thermo Electron Jouan C3i multifunction centrifuge for 5 minutes and excess media was decanted off. The cell pellet was washed twice with cold PBS and the cell pellet was either stored at -80°C until required or lysed immediately.

Cell lysis was performed as per Section 2.6.1 prior to performing Filter-Aided Sample Preparation (FASP). FASP which had the benefit of aiding detergent removal and desalting on a single filter unit.

2.6.3.7. Filter Aided Sample Preparation

All centrifugation steps took place in a Hettich Mikro200R centrifuge at 14,000Xg for 40 minutes at 20°C unless otherwise stated

Sample Preparation

100µg of protein was denatured by the addition of 1µL of 0.5M DTT (prepared in 50mM Ammonium Bicarbonate) to each sample followed by incubation at 56°C in a dry heat block for 20 minutes. This was allowed to cool and 200µL of 8M Urea was added to each sample prior to transferring to a Microcon centrifugal filter unit (Merck Millipore, MRCPRT010) with a 10kDa molecular weight cut off. The samples were centrifuged, the filtrate discarded and 200µL of wash buffer 1 (8M Urea, 0.1M Tris-HCl pH8.5) added to the retentate. The samples were centrifuged, the filtrate

discarded and 100µL of 0.05M Iodoacetamide (prepared in wash buffer 1) added to the retentate. The samples were shaken at 600rpm for 60 seconds and incubated in the dark for 15 minutes to alkylate them. The samples were centrifuged, the filtrate discarded and 100µL of wash buffer 2 (8M Urea, 0.1M Tris-HCl pH7.9) added to the retentate. The samples were centrifuged, the filtrate discarded and a further 100µL of wash buffer 2 (8M Urea, 0.1M Tris-HCl pH7.9) was added to the retentate and the samples centrifuged.

Protein Digestion with Trypsin

Prior to digesting with MS grade Pierce™ Trypsin Protease (Thermo Fisher Scientific, 90057) the permeate was diluted with 100µL 50mM Ammonium Bicarbonate. The trypsin was prepared by reconstituting lyophilised trypsin in 50mM Acetic Acid to a final concentration of 1mg/mL. 2µL Trypsin (1mg/mL) was added to each sample and the tubes sealed shut with parafilm. Samples were digested overnight at 37°C in a dry heat block at 300rpm. Following digestion, samples were centrifuged, and the filtrate retained. The Microcon centrifugal filter unit was then rinsed with 50µL of 0.5M NaCI and centrifuged at 14,000Xg for 20 minutes to wash any residual peptides from the walls of the filter unit into the filtrate. The filtrates were acidified with Acetonitrile (ACN) and Trifluoroacetic acid (TFA) to a final concentration of 2% ACN, 0.1% TFA and stored at -80°C.

C18 Clean up and Concentration Normalisation

Pierce™ C18 spin columns (Thermo Fisher Scientific, 89870) containing 8µg C18 resin with a binding capacity of 30µg total peptide were used to normalise the peptide concentration between samples prior to loading on the MS for analysis.

All reagents were prepared in MS grade H_2O unless otherwise stated. All centrifugation steps used a Hettich MikroR200 at 1500Xg for 1 minute at 20°C unless otherwise stated.

Each column was tapped to ensure all the resin was at the bottom of the column prior to activation with 200µL of a 50:50 ACN:H_2O solution, followed by centrifugation. The flow through was discarded and the columns equilibrated with 200µL of a 2%:0.1% ACN:TFA solution, followed by centrifugation. The acidified peptide sample (from protein digestion) was applied to the top of the C18 column, and the column centrifuged. The column flow through was recovered and applied

to the C18 column twice more to ensure sufficient binding of the peptides to the C18 resin. The C18 column was washed twice with 200µL of a 2%:0.1% ACN:TFA solution and the flow through discarded after each wash. Peptides were eluted using two sequential centrifugation steps with 20µL 70:30 ACN:H_2O each time. The flow through was retained after each step and combined. The solvent was evaporated to dryness in a Jouan Maxidry plus SpeedVac vacuum evaporator. The dried peptides were resuspended in 20µL of a 2%:0.1% ACN:TFA solution. The peptide concentration of each sample post resuspension was quantified using a Nanodrop One spectrophotometer (Thermo Scientific) using the Scopes method at 205nm (Goldfarb et al., 1951). The loop on the LC-MS in use at the NICB has a 6.4µL sample volume so samples were further diluted with a 2%:0.1% ACN:TFA solution to 1µg/3µL if appropriate. If this was not possible due to samples being too dilute, samples were either dried down further so this could be achieved or 5µL of sample was loaded on the MS and the absolute amount of peptide loaded noted.

2.6.4. MS/MS Operating Conditions

The samples for MS were run by Dr. Michael Henry of the NICB

Reverse-phased capillary high pressure liquid chromatography was performed using an UltiMate 3000 nanoRSLC LC system (Thermo Scientific) coupled directly in-line with the Thermo Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific).

The digested protein samples (1µg peptide) were loaded onto the trapping column (PepMap100, C18, 300µm x 5mm) at a flow rate of 25µl/min with 2% (v/v) acetonitrile (ACN), 0.1% (v/v) trifluoroacetic acid (TFA) for 3 minutes before being resolved onto an analytical column (Acclaim PepMap 100, 75µm x 50cm, 3µm bead diameter column). Peptides were eluted using the following binary gradient; solvent A (0.1% (v/v) formic acid in LC-MS grade water) and solvent B (80% (v/v) ACN, 0.08% (v/v) formic acid in LC-MS grade water) using 2-32% B for 75 minutes, 32-90% B in 5 minutes and holding at 90% for 5 minutes at a flow rate of 300nl/min. For peptide ionization, a voltage of 2.0kV was applied and a capillary temperature of 320°C was used. Data-dependent acquisition with full scans in the 380-1500 m/z range was performed using an Orbitrap mass analyser with a resolution of 120,000 (at m/z 200), a targeted automatic gain control (AGC) value of 4E+05 and a maximum injection time of 50ms. The number of selected precursor ions for

fragmentation was determined by the top-speed acquisition algorithm. Selected precursor ions were isolated in the Quadrupole with an isolation width of 1.6Da. Peptides with a charge state of 2+ to 7+ were analysed and a dynamic exclusion was applied after 60 seconds. Precursor ions were fragmented using higher energy collision-induced dissociation (HCD) with a normalized collision energy of 28%, and resulting MS/MS ions were measured in the linear ion trap. The typical MS/MS scan conditions were as follows: a targeted AGC value of 2E+04 and a maximum fill time of 35ms.

2.6.5. Differential Protein Analysis using Quantitative label-free LC-MS/MS Proteomics

Two software packages were used to analyse the LC-MS/MS raw data files and generate lists of proteins whose expression level differs between experimental conditions.

> Differential protein expression (DE) analysis was performed using the software package Progenesis QI for Proteomics, version 2.0 from Non Linear Dynamics (a subsidiary of Waters).

> Proteome Discoverer, version 2.2 (Thermo Fisher Scientific) in conjunction with the SEQUEST (HT) and Percolator algorithms and a proteogenomic draft annotation of the Chinese hamster *(Cricetulus griseus)* genome (Li et al., 2018). was used to identify individual peptides and the proteins they came from.

Progenesis QI for Proteomics

Raw MS data files were imported into the Progenesis QI for Proteomics interface and automatic run alignment was performed to minimise variability between samples and replicates. The alignment scores of the individual replicates for each sample were reviewed.

Following alignment, samples were split into experimental groups and identified features were filtered based on an ANOVA p value of < 0.05 between experimental groups. A mascot generic file (.mgf) was generated for each replicate and exported to Proteome Discover. The SEQUEST (HT) and Percolator algorithms were used to identify individual peptide sequences and the proteins they were derived from. The parameters utilised were as follows:

- MSI tolerance =10ppm
- MS2 mass tolerance = 0.6Da
- Enzyme specificity was set as trypsin with a maximum of two missed cleavages allowed.
- Allowed modifications were oxidation and carbamidomethylation
- Data was filtered against a 1% false discovery rate (FDR) on peptide spectral matches (PSM) using automatic decoy searching in SEQUEST (HT)
- PSM were filtered using an Xcorr value of ≥ 1.5
- Peak lists were searched against a proteogenomic draft annotation for the Chinese hamster genome (Li et al., 2018).

Following peptide and protein identification by Proteome Discover the data was saved as a (.pepxml) file and reimported into Progenesis QI for Proteomics. Criteria of an ANOVA p value of < 0.05, a relative minimum fold change of 1.2 between experimental groups and ≥2 peptides contributing to protein identifications was used to identify proteins whose expression profile differed between experimental groups.

Differential Expression Analysis and Statistics

As part the of identification of DE proteins an FDR filter (q value ≤ 0.01) was applied using false decoy searching via the Percolator module in Proteome Discoverer. It was also applied during the identification of significantly enriched KEGG pathways and gene ontology terms (q value ≤0.05). However, FDR was not employed during the identification of significantly DE peptides using Progenesis QI for Proteomics.

This thesis makes use of the principle of category-based inductive reasoning, i.e., the ability to make true predictions based on a set of empirical observations. This approach is based upon the assumption that because gene or protein X has been shown to have a particular function in humans, it will have the same function in all mammals. Hence why *Homo sapiens* and not *Cricetulus griseus* has been employed as the reference data set in the bioinformatics analysis employed in this thesis.

One of the advantages of employing a MS-MS strategy in the analysis/comparison of protein expression between experimental groups is the vast number of ion peaks that

are reported in each group. Each of these individual ion peaks can be matched to a peptide sequence and the intensity of the ion peak is related to that peptide's abundance within the sample. Differential expression analysis compares each peptide's abundance from the same sample in each experimental group with the aim of identifying peptides which differ in their intensity/abundance between the experimental groups. Lualdi and Fasano (2019) identified two issues that impact proteomics datasets such as the one discussed in this thesis . These are dimensionality and sparsity. Dimensionality describes the number of variables (peptides) identified. In data sets with high dimensionality the number of features identified greatly exceeds the number of samples tested, e.g., in chapters 5, 6 and 7 thousands of peptides were identified but only two experimental groups were examined for each cell line. It stands to reason that in datasets with high dimensionality the chance of a false positive increases and hence the requirement for a statistical filter to ensure this doesn't happen. Sparsity, a low number of replicates for each experimental condition is also often associated with proteomic datasets. Together these two issues can lead to over fitting whereby small differences in ion intensity are incorrectly associated with a statistically significant difference in peptide abundance.

Therefore, an assessment should be made as to whether differences in protein abundance were identified by chance or represent a true difference. Standard practice is to assess this by applying a statistical cut off (p value ≤0.05) to the data. In this study an ANOVA was used to identify peptides and proteins whose abundance differed significantly between experimental groups. While this type of approach is the standard practice when comparing experimental groups which differ in a single variable the analysis of experimental groups which contain multiple variables requires the application an additional statistical filter called the false discovery rate (FDR). It's beyond the scope of this thesis to discuss the theory behind FDR and the different approaches that can be employed when applying it to proteomics data, see Jung (2016) and Lualdi and Fasano (2019) for reviews. However, it's important understand the impact of omitting an FDR filter during the identification of peptides which differed significantly between experimental groups has on the subsequent data analysis.

A p value ≤0.05 means that for every one hundred repetitions of a given experiment the observed result should occur by random chance on five or less occasions and hence

provides confidence that the observed result is the true result. However, while this holds true when only one variable is being assessed, it doesn't hold true for multi variate analysis. In multi variate analysis (such as proteomics) where multiple variables are being assessed simultaneously, applying a p value of ≤ 0.05 at the individual peptide level provides confidence that that peptide differs in its abundance between experimental groups. However, it doesn't provide confidence that all peptides selected will have a collective p value of ≤ 0.05 as the possibility remains that all the peptides selected in this way represent type 1 errors or false positives, i.e., the null hypothesis was rejected when it should have been accepted. The application of an FDR filter of ≤ 0.01 would have provided confidence that the probability of the entire set of peptides identified as being differentially expressed in this study represented a false positive was ≤ 1%.

Given an FDR filter was not applied during the identification of DE peptides it's appropriate to consider the implications this may have on the list of DE proteins identified in this study. No FDR filter, no matter how stringent it is, can guarantee that the entire set of DE peptides don't represent a false positive. Additionally, as discussed by Lualdi and Fasano in their 2019 review "the problem of false positives when performing multiple statistical tests (i.e., several variables are considered simultaneously) is managed by the application of multiple testing corrections (MTCs). However, this cannot be done without affecting the probability of type II errors, which concomitantly increases." Type II errors are false negatives, i.e., the null hypothesis being accepted when it should have been rejected. In this study, the chances of a false positive at the DE peptide level leading to a false positive at the DE protein level were reduced by:

- only identifying a protein as being DE if at least two unique peptides had been identified as being associated with it.
- Use of an additional filter based on the Xcorr value assigned to that protein by the Proteome Discoverer package. Xcorr values measure how well the experimentally determined spectra for a peptide X matches the theoretical spectra generated for that peptide under the parameters chosen for that particular experiment. An Xcorr value of 1.5 as reported by Henry et al. (2017) and Kaushik et al. (2018) was applied.

- Applying an FDR rate of 0.01 (1%) when using the Proteome Discoverer package, i.e., at the protein level (using the Percolator algorithm).

Proteome Discoverer uses the decoy proteome approach devised by Peng et al. (2002) to estimate FDR rates. Briefly, this approach involves searching each DE peptide against composite database which consists of a protein sequence (original) database and its reversed version. Using this approach, the FDR is calculated as follows:

$$FDR = 2*n(rev)/(n(rev)+n(forw))$$

where nrev and nforw are the number of number of peptides using the reversed and forward (original) databases respectively.

For chapter 5 the data was analysed and reported using a 1% FDR filter at the DE protein level in both Proteome Discoverer and Progenesis QI for Proteomics. For chapters 6 and 7 the data was analysed using a 1% FDR filter at the DE protein level in Proteome Discoverer only and an assessement of what impact that had on the reported results was made.

2.6.6. R Packages Used in the Analysis of DE Protein Lists

The R statistical framework (R Core Team (2021) (R v4.1.1, www.R-project.org) and the Bioconductor (Morgan, 2021) repository (www.bioconductor.org) includes multiple packages for identifying enriched genes and metabolic pathways. The ones used in this project include:

- org.Hs.eg.db (Carlson, 2021) version 3.13.0 - genome wide annotation for *Homo sapiens.*
- AnnotationHub (Morgan and Shepard, 2021) version 3.0.2 - used for performing enriched GO analysis
- biomaRt (Durink et al., 2005) version 2.48.3 - interface to Ensembl database
- clusterProfiler (Yu et al., 2012) version 4.0.5 - enrichment tool for analysing omics data
- GO.db version 3.13.0 - annotation maps describing Gene Ontology
- mygene (Mark et al., 2021) version 1.28.0 - used to match gene identifiers to protein names and Uniprot ID.

- pathfindR (Ulgen et al., 2019) version 1.6.3 - package for performing Peptide Set Enrichment Analysis (PSEA)
- rrvgo (Sayols, 2021), version 1.4.4 - package for reducing the number of GO terms by identifying redundancy based on semantic similarity
- NMF (Gaujoux and Seoighe, 2010), version 0.23.0 - package used to draw heatmaps
- Figures were drawn with the use of ggplot2 (Wickham, 2016), version 3.3.5, UpsetR (Conway et al., 2017) version 1.4.0) and EnhancedVolcano (Blighe et al., 2021), version 1.10.0 packages.

Scripts are available upon request.

2.6.7. Unsupervised Hierarchical Clustering

Unsupervised hierarchical clustering (heatmaps) was performed using the aheatmap function within the R statistical framework as a quality control step. The Euclidean distance correlation was applied to normalised abundance data from Progenesiss QI for Proteomics.

2.6.8. Principal Component Analysis

PCA is a multivariate analysis that aims to explain the variance of the data by transforming it to a new coordinate system (each coordinate called a principal component) usually after a normalisation step. This allows the maximum amount of information in a dataset to be described using the least possible number of variables (principal components). It's standard practice for principal component 1 to describe the largest amount of variation in the data and sufficient components are assigned until all the variation is described. The PCA model describing the variation in the data is considered a good fit if most of the variation in the dataset is described by principal components 1 and 2. This simpler picture can help assess the quality of the data (i.e., replicate samples should cluster together) and can help in the interpretation of the data by matching variables with components. It has as its aim increasing ease of interpretation while minimising information loss. In an ideal scenario, all the variation in a dataset would be described by one variable (principal component 1), but this rarely (if ever) occurs. The specific PCA model employed by the Progenesis QI for Proteomics

software used the abundance level of each individual DE protein to determine the variation in the data and therefore experimental groups and replicates can be separated based on differences in the abundance of each protein. This aids the identification of outliers.

2.6.9. Gene Ontology Analysis (GO)

Gene Ontology (GO) annotation had to be inferred from human assignments. Systematic GO assignments are available for CHO but using human ones resulted in an increased number of enriched terms be identified. The Bioconductor packages biomaRt, AnnotationHub and clusterProfiler, were used to assign HGNC symbols to the differentially expressed proteins, perform Gene Ontology (GO) analysis with respect to cellular location and biological process of the differentially expressed genes and identify significantly enriched GO terms and biological processes. The criteria for calling a term significantly enriched were p-adjusted (Benjamini-Hochberg) and q-value cut-offs < 0.05. The Bioconductor package rrvgo was used to perform binning of enriched child GO terms to their parental terms. This approach reduced the list of enriched GO terms, and consistent biological themes were identified.

2.6.10. KEGG Pathway Profiling and PSEA

The differentially expressed proteins were assigned to KEGG pathways (Kanehisa et al., 2014 and Kanehisa and Goto, 2000) with the help of the pathfindR package. The criteria for calling a pathway significantly enriched were p-adjusted and q-value cut-offs < 0.05.

The benefit of this R package, over other packages, was that it also performed protein set enrichment analysis which allowed a status of either upregulated or downregulated to be assigned to each differentially expressed proteins. Additionally, it provided pathway diagrams of the enriched pathways.

2.6.11. Volcano Plots

Volcano plots were a method of visually displaying large datasets so that statistical significance and both the degree of change from the control and direction of change are presented in one plot. It involved calculating the fold change by dividing the average values obtained for each protein in the test condition (e.g., no serum) by the average values obtained for the same protein in the control condition (e.g., serum).

The $\log_2$ of thefold change was calculated and plotted on the x-axis. The negative $\log_{10}$ of the p value obtained from an ANOVA for each protein was plotted on the y-axis.

2.6.12. Upset Plots

Upset plots are variations on a Venn diagram that are useful for displaying the overlaps between multiple datasets. The technique was first reported by Lex et al. (2014).

2.7. Flocculation Experiments Using Cations

2.7.1. Cell Culture

Balan CD was seeded with CHO K1 cells were routinely cultured in 250mL Corning® shake flasks (50mL working volume) in a Kühner orbital shaker at 37°C, 5% CO_2, 80% RH and 170rpm. BalanCD® (Fujifilm) supplemented with 4mM L Glutamine (Gibco) was seeded with CHO K1 cells at $5X10^6$ cells/mL. The cells were allowed to grow until the cell density was in excess of $1X10^7$ cells/mL. All experiments were carried out at either $2X10^7$ cells/mL or $8X10^6$ cells/mL in spent or fresh media. Routine cell viability and cell density measurements were performed using the ViaCount™ assay on a Guava® EasyCyte™ benchtop flow cytometer (Luminex).

2.7.2. Preparation of Cation Stocks

5M stocks of $CaCl_2$, $MnCl_2$ and $FeCl_3$ were prepared in sterile UHP water and filtered with a 0.22µm filter (Millipore, Millex-GV, SLGV-025BS). 10X working solutionsin sterile UHP water of each cation were prepared from the 5M stock solution as required.

2.7.3. Assessing the Effect of pH on Cellular Flocculation on the Flocculation

Overnight cultures of CHO K1 cells were counted, centrifuged, and resuspended in a volume of either fresh or spent media (Section 2.7.4) such that the density of the resulting cell suspension was either $2X10^7$ cells/mL or $8X10^6$ cells/mL depending on the experimental design. 4.5mL of cell suspension was transferred to a TPP 50mL TubeSpin® bioreactor and a 0.5mL of a 10X solution of the cation under investigation was added so that the final cation concentration was 1X. Cells and cation were incubated in a TPP 50mL TubeSpin® bioreactor and sampled at 15

minutes and 60 minutes post addition. After each timepoint an aliquot of cells was removed and cell viability, cell concentration, cell diameter and %aggregation analysis of flocculated CHO K1 cells was performed using the Trypan Blue exclusion method (Section 2.3.1) and a Cedex cell counter.

2.7.4. Assessing the Effect of Cation Concentration on the Flocculation of CHO K1 cells

The effect of culture pH on the ability of cells to flocculate was investigated by resuspending CHO K1 cells in either "spent" media from overnight cultures or in "fresh" media. The theory behind this was that the cells in spent media would have a lower pH than those in fresh media.

Cells were grown, resuspended, cation added, and analysis performed as per Section 2.7.3.

The results of the experiments described in sections 2.7.3 and 2.7.4 determined the cation concentrations used in section 2.8

2.8. Assessing the Effect on Product Quality Following Flocculation of CHO K1 Cells with Divalent Cations

2.8.1. Cell Culture

CHO K1 cells were grown and resuspended in spent media at $2X10^7$ cells/mL as per Section 2.7.3. Human IgG (Abcam, ab91102) was then added to the flask of cells and allowed to mix at 37°C, 170rpm in a ISF1-X (Climo Shaker) Kühner cell culture incubator for fifteen minutes. 4.5mL of cells was transferred to TPP 50mL TubeSpin® bioreactor and 0.5mL of a 10X cation stock was added as per Table 2.4. Cells were returned to the incubator and allowed to mix at 37°C, 170rpm incubator for fifteen minutes. 0.5mL of cells were removed and 0.5mL 0.5M EDTA was added. Cells were returned to the incubator and left standing without shaking at 37°C, 0 rpm in a ISF1-X (Climo Shaker) Kühner cell culture incubator for ten minutes while sedimentation occurred. Sampling time points are detailed in Table 2.4 and analysis was performed as per Section 2.7.3.

Experimental Condition	10X Cation stock	Sampling Time Points
Negative Control	PBS	Immediately prior to cation addition
		Immediately prior to EDTA addition
		Post Flocculation/Sedimentation
Test	1.0M $CaCl_2$	Immediately prior to cation addition
		Immediately prior to EDTA addition
		Post Flocculation/Sedimentation
Test	0.5M $MnCl_2$	Immediately prior to cation addition
		Immediately prior to EDTA addition
		Post Flocculation/Sedimentation
Test	0.25M $FeCl_3$	Immediately prior to cation addition
		Immediately prior to EDTA addition
		Post Flocculation/Sedimentation

Table 2.4: Cation Concentrations Used in Section 2.8.1

2.8.2. Measurement of Lactate Dehydrogenase Activity in Cell Culture Supernatant

The level of lactate dehydrogenase activity in CHO cell culture supernatants was used as an orthogonal method for measuring cell viability and assessing the degree of cell lysis which maybe occurring due to the addition of highly concentrated cation solutions. The Lactate Dehydrogenase Activity Assay Kit (Sigma, MAK066) was used in accordance with the manufacturer's instructions as follows. Cell pellets were removed from storage at -80°C and homogenised on ice with 500µL of cold LDH assay buffer. Insoluble material was removed by centrifugation in a Sigma 1- 1EPK

bench top centrifuge for 15 minutes at 4°C and 10,000Xg. Samples and NADH standards were diluted to a final volume of 50µL with assay buffer and added to a 96 well plate. LDH substrate solution was prepared by diluting the LDH substrate mix in assay buffer (2µL substrate mix and 48µL assay buffer/well) and 50µL was added to each well containing samples and standards and blanks. The plate was incubated at 37°C and the absorbance measured at 450nm, after 3 minutes incubation and at 5 minute intervals thereafter until the highest sample absorbance exceeded the absorbance of the top standard. The blank was subtracted from all readings prior to calculating the level of LDH activity in the samples.

2.8.3. Protein A/G Purification of Antibody from Flocculated Cells

NAb™ Columns containing 1mL of recombinant Protein A/G fusion protein immobilised to a 6% crosslinked beaded agarose (CL-6B) resin (Thermo Fisher Scientific, Product No 89958) were used to purify antibody from the cell free supernatant from each sampling time point described in section 2.8.1. Cells were pelleted immediately after sampling by centrifugation at max speed in an Eppendorf 5424 microfuge for 5 minutes. Cell free supernatant was removed to a clean 1.5mL microcentrifuge tube (Eppendorf) and stored at -80°C until required for analysis. The manufacturer's protocol was modified as follows to allow for purification of the entire supernatant volume. Cell free supernatant was passed over the resin by spinning the column in a centrifuge and the flow through collected. All centrifuge steps were performed in a Sorvall Legend T benchtop centrifuge at 1000Xg for 1min at room temperature.

The purification protocol involved five steps:

- Storage buffer was removed from the NAb™ recombnant Protein A/G fusion protein columns by centrifuging the columns.
- The resin was then equilibrated passing at least 4 column volumes of binding buffer through it.
- Sample was mixed with an equal volume of binding buffer (Thermo Fisher Scientific Product No 28372). 2mL of diluted sample was applied to the resin, the column capped and incubated on a rotary mixer for 10 min to allow any antibody present in the sample bind to the resin. The column was uncapped, and sample removed by centrifugation. This step was repeated as often as

necessary to apply the entire sample to the column.

- Unbound material was removed from the column using three sequential 2mL washes with binding buffer.
- Finally bound antibody was eluted from the resin using 3 sequential 1mL washes with IgG Elution buffer, pH 2.8 (Thermo Fisher Scientific Product No 1856202).

The amount of antibody present at each step was quantified using an ELISA assay specific for human IgG.

2.8.4. Size Exclusion HPLC of Purified IgG

Purified samples (100µL, Section 2.8.3) were analysed on an Agilent 1100 HPLC equipped with a variable wavelength detector for detection at 280nm using a TOSOH TSK-GEL® guard column (TOSOH, 08543) and G3000SWxi 30cm column resolving column with a 7.8mm internal diameter (TOSOH, 08541). The components present in the purified samples (Section 2.8.3) were resolved using an isocratic elution with 0.22µm filtered 150mM phosphate buffer pH 7.0 at 1.0mL/min. Analysis was performed using Chemstation software as described by Schrag et al. (2014). The output of an SEC-HPLC analysis is a plot of absorbance @280nm (mAU) versus time (minutes). The relative amount (%) of monomer, aggregates and fragments were determined by integration of the area under the curve for each peak which corresponded to these species and expressing that area as a percentage of the total integrated peak area. Molecular weights were assigned to the various components present in each sample based on retention time by reference to standards of known molecular weight (Bio-Rad, 1511901).

2.9. Enzyme Linked Immunosorbent Assay (ELISA) for Antibody Quantification

The concentration of human IgG was monitored throughout the spiking and recovery experiments using the Bethyl Laboratories Human IgG ELISA Quantitation Set (Bethyl Labs Inc., E80-104) in accordance with the manufacturer's instructions. Coating Buffer (0.05M Carbonate-Bicarbonate, pH9.6), Blocking buffer (TBS, 1% BSA, pH8.0), Wash buffer (TBS, 0.05% Tween 20, pH8.0), Sample Diluent (TBS, 1% BSA, 0.05% Tween 20, pH 8.0) and Stop Solution (0.18M Phosphoric acid) were prepared in accordance with the manufacturer's instructions. Standards, blanks and

samples were analysed in duplicate. A 96 well ELISA plate was coated with 100µL coating antibody (1µL diluted in 100µL of 0.05M Carbonate-Bicarbonate Buffer, pH9.6 per well coated) and incubated at 4 °C overnight. The plate was washed five times with 100µL wash solution (TBS - 0.05% Tween 20, pH8.0) and the each well incubated with 200µL blocking buffer (TBS - 1% BSA, pH8.0) for between 1 - 2 hours at RT to prevent non-specific binding. During this incubation step the standard curve was prepared and samples were diluted based on their expected concentration so that they would fall on the linear portion of the standard curve. Samples and standards were diluted in TBS, 1% BSA, 0.05% Tween, (pH 8.0). Standards were prepared from the same IgG that was used to spike the CHO cell cultures. A blank consisting of the standard and sample diluent was included in duplicate in all assays. After incubation with blocking buffer the plate was emptied, and the wells were washed five times as previously described. 100µL of each standard, blank and sample was added to the plate in duplicate and the plate was incubated for 1 hour at RT. The HRP conjugated detection antibody was diluted 1/100,000 in TBS, 1% BSA, 0.05% Tween, (pH 8.0) prior to use. After incubation with standards, samples and blanks the plate was emptied and each well washed five times as previously described. 100µL of diluted HRP-detection antibody was added to each well and the plate incubated at RT for 1 hour. The TMB substrate was allowed to equilibrate to RT during this period. After incubation with HRP-detection antibody the plate was emptied, and the wells were washed five times as previously described. 100µL of TMB substrate solution (Sigma Aldrich T8665-100mL) was added to each well and the plate incubated at RT in the dark for 10-15 minutes. The presence of Human IgG was confirmed by a colour change from colourless to blue. After 10- 15 minutes the enzymatic reaction was stopped by the addition of 100µL 0.18M Phosphoric Acid (Sigma Aldrich, 04102) to each well, which resulted in a colour change from blue to yellow. The absorbance of each well @450nm was read on a Multiskan™GO Microplate Spectrometer and a standard curve of absorbance at 450nm against IgG standard concentration was constructed. This allowed the amount of IgG in each sample to be determined once dilution factor was accounted for.

CHAPTER 3

Identification, Cloning and Expression of Candidate Genes for Promoting Flocculation of CHO cells

3.1. Introduction

While biological therapeutic molecules have been produced in a wide variety of cells (reviewed by Lalonde and Durocher, (2017), CHO cells are the mainstay of the biologics industry, accounting for over 50% of all licensed molecules (Dumont et al., 2016). A recent report in Bioprocess International (Stanton, 2021) estimated that by 2025 annual global demand (including both commercial and clinical requirements) for these molecules will approach 64,000 Kg, up from 32,000 Kg in 2020.

One of the quickest and simplest ways of increasing recombinant product yield per unit volume is to increase the cell concentration and viability in the manufacturing vessel. A fed batch STR process capable, at harvest, of supporting viable cell concentrations of $4X10^7$ cells/mL, with a culture viability greater than 90% has been reported (Dr Andy Racher Lonza Biologics, personal communication). However, this concentration of cells can lead to problems during primary recovery when these types of process are scaled up to the volume required to support market supply. Consequently, the increase in overall yield (g recombinant protein/unit volume) which resulted from increases in product titre may not be realised. This is because increases in cell concentration of this magnitude will adversely impact the efficiency of disk stack centrifuges and filters in clarifying the bioreactor contents. These are the two most common techniques used by industry to clarify the bioreactors contents. Consequently, the yield from primary recovery operations could be reduced.

3.1.1. Flocculation as a solution to the primary recovery challenges associated with high viable cell concentration at harvest.

A solution to the primary recovery challenges associated with increased cell concentration and high pcv values could be flocculation. From a bioprocess perspective flocculation involves the cells coming out of suspension and forming flocs or aggregates. CHO cells don't spontaneously flocculate so flocculating agents such as polydiallyldimethylammonium chloride (pDADMAC), are typically added to the bioreactor broth (McNerney et al., 2015).

One of the potential issues with the use of flocculating agents, such as pDADMAC, in biologics production processes, is that regulators require evidence that the purification process can remove all traces of the flocculating agent from the final purified

recombinant protein. This would potentially involve the development and validation of clearance assays specific to the flocculating agent, which is an expensive and time-consuming task. This work examined a potential solution to this issue. This chapter describes an investigation into the feasibility of expressing an exogenous protein, which could promote cellular flocculation, into CHO cells. Induction of this protein's expression prior to harvest would result in flocculation of the bioreactor contents. Examples of such proteins are the *flo* family of genes in *S. cerevisiae*. A literature search also identified, Hydrophobins, a family of fungal proteins and a mammalian protein, Aquaporin-0, as candidates for this role. Accordingly, ABH1, HFB1, (Hydrophobins from *Agaricus. bisporus* and *Trichoderma reesei* respectively), Flo5 and Flo11 and Aquaporin 0 (AQP0) were selected as the proteins to be studied in this chapter.

3.2. Results

3.2.1. Synthesis of Gene Sequences

The gene sequences for all five proteins were obtained from GenBank (Appendix 1 for the sequences and GenBank information), checked for the presence of restriction enzyme sites that could interfere with subsequent cloning and their anticipated subcellular expression location confirmed by Protter analysis (http://wlab.ethz.ch/protter/start/) (Appendix 2). Prior to submitting these sequences to Genscript (a commercial vendor) for synthesis some additional restriction enzyme sequences were added upstream of the ATG start codon and downstream of the stop codon to aid subsequent recloning, a KOZAK sequence (GCCGCCACCATGGGG) was added upstream of the N terminal ATG and a 6X His label was added to the 3' end of each gene (C terminus of the protein) to aid detection by western blot.

The sequence for the eGFP gene was obtained and processed in an equivalent manner prior to being synthesised by Genscript. This construct did not have a 6X His label. This is because the transfection protocol followed utilised co-transfection of two plasmids. One plasmid contained the sequence for the protein of interest and the other plasmid contained the sequence for eGFP. This allowed for transfection efficiency to be assessed using the degree of fluorescence of the cells due to expression of eGFP. However, the fact that the transfected cells expressed both eGFP and the cloned protein means that if both were His tagged there would be two bands detected on western blots. This would

not be an issue for lysates from cells transfected with Hydrophobin genes but would complicate detection on the western blots for AQP0 as they have similar molecular weights.

The synthesised genes were provided in the pcDNA3.1 expression vector with Hygromycin as the selectable marker.

3.2.2. Transgene Expression in Transiently Transfected CHO S Cells

One mL of CHO S cells @1X10^6 cells/mL were transfected with 500ng total DNA using Freestyle transfection reagent as per the methods section (Section 2.4.6.2). Transient transfections were performed in twenty-four well plates and allowed to grow for twenty-four hours before 500µL of cell suspension was removed for RNA isolation. The remaining cell suspension was refed with fresh pre warmed media and returned to the incubator for a further 72 hours to allow for protein expression.

DNase treated RNA, isolated from transiently transfected CHO S cells (Section 2.4.7), was used as the template in a cDNA synthesis reaction with random primers and reverse transcriptase (Section 2.4.8.2). This was subsequently used as the template in a PCR reaction (Section 2.4.9) with primers specific for GAPDH, ABH1, HFB1 and AQP0 respectively (Figure 3.1). A cDNA synthesis reaction excluding the enzyme reverse transcriptase was used as a negative control for each reaction.

Raw data (Ct plots) are presented due to the difficulties of calculating RQ because the target genes were not exogenously expressed in CHO cells and initial experiments did not include untransfected negative controls. This made normalisation difficult to achieve. Additionally, the purpose of performing these experiments was to simply confirm that the transfected genes were being transcribed, not to quantify the amount of transcription occurring.

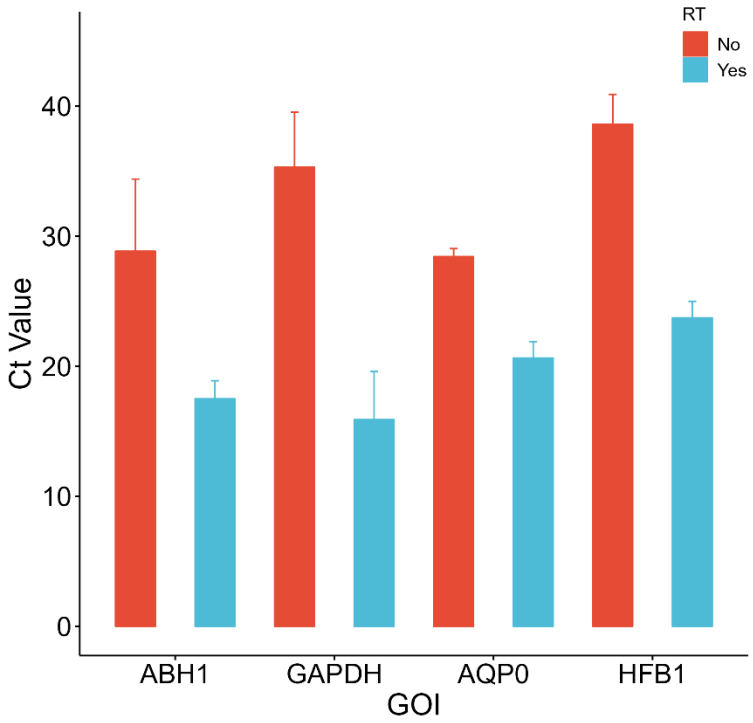


Figure 3.1: **Average Ct values obtained for the target genes following RNA isolation from transfected CHO S cells and subsequent cDNA synthesis with and without reverse transcriptase. The bars are the average of three biological replicates and each biological replicate was itself the average of three technical replicates.**

For all genes, the difference between template generated with and without reverse transcriptase ranged from 7.8 cycles or 2.4 $\log_{10}$ (3.3 cycles in a RT-PCR reaction is a 10-fold difference in concentration) for AQP0 to 19.9 cycles or 6.0 $\log_{10}$ for GAPDH. A Ct (critical threshold) value ≥ 35.0 is generally accepted as being indicative of the fact that no template was detected in the sample. None of the wells containing template generated with reverse transcriptase exceeded the cut-off threshold of thirty-five cycles. This indicates that each of the four genes which were transfected into CHO S cells maybe being transcribed into mRNA by the CHO S cells. The reason the last sentence is qualified with a maybe is because of the Ct values obtained for template generated without reverse transcriptase. The absence of reverse transcriptase from the cDNA synthesis reaction means that cDNA can't be present in that sample. Accordingly, all wells which represented template generated without reverse transcriptase should have a Ct value ≥ thirty-five. This threshold was only met for GAPDH (35.3) and HFB1 (38.6).

An examination of the Ct values for each individual biological replicate suggested that for one of the rounds of transfection of ABH1, an error was made during template

synthesis or RT-PCR sample preparation. The average Ct value for ABH1 without reverse transcriptase in this round was 22.7 while the average Ct value for ABH1 with reverse transcriptase in this round was 16.8. These Ct values suggest that the target sequence for ABH1 was detected regardless of the presence or absence of reverse transcriptase during the cDNA synthesis step. Once this round was excluded the average Ct value for the three biological replicates of ABH1 without reverse transcriptase increased from 28.8 to 33.0.

No such error was identified for the AQP0 gene. This left three explanations for the signal obtained for AQP0 in the absence of reverse transcriptase:

- Primer dimers were being formed, amplified, and detected.
- The RNA preparations used in the cDNA synthesis reaction were contaminated with DNA and the DNase treatment performed as part of this procedure was not sufficient to remove this.
- The primer set used in the RT-PCR reaction recognised sequences in the CHO S genome and what was being detected in the RT-PCR reaction was CHO AQP0 cDNA not human AQP0 cDNA.

In an attempt, to answer this question RNA was isolated from untransfected CHO S cells and used to generate cDNA in the presence and absence of reverse transcriptase. This cDNA was used as the template in a RT-PCR experiment with the primer set for each of the transfected genes. The rationale was, if the Ct value observed for the ABH1, HFB1 and AQP0 genes was solely due to nonspecific amplification using the same primer set with transfected and untransfected cells would result in the same Ct value being obtained for transfected and untransfected cells when reverse transcriptase was included in the cDNA synthesis step. The results are presented in Figure 3.2.

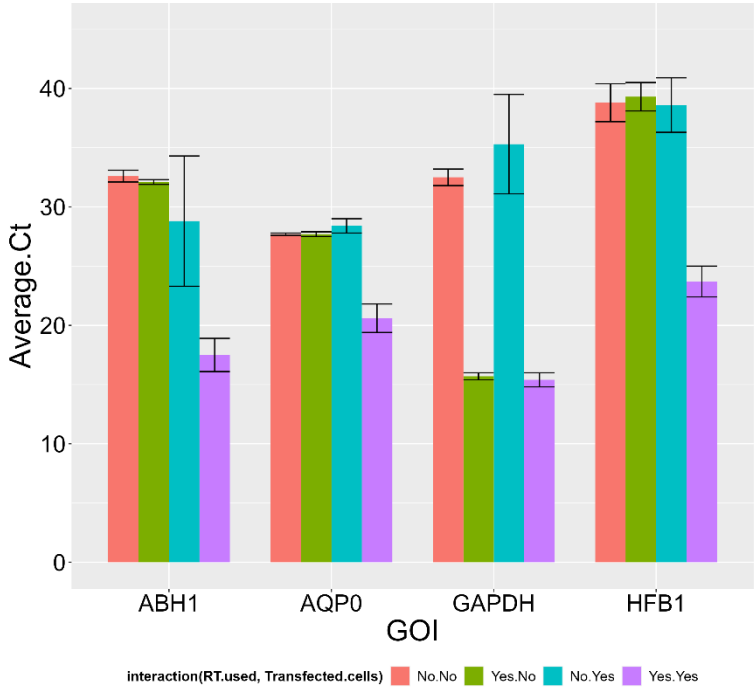


Figure 3.2: **Average (3 biological replicates) Ct values obtained for target genes following RNA isolation from transfected and untransfected CHO S cells and subsequent cDNA synthesis (including and excluding reverse transcriptase). The error bars represent three technical replicates.**

Whether transfected or untransfected cells were used as the RNA source didn't matter in the case of GAPDH. Once reverse transcriptase was included in the cDNA synthesis reaction a signal was observed. This was as expected. GAPDH was chosen as the endogenous (positive) control. The primer set used was designed to detect transcription of CHO GAPDH gene. Differences were observed in the Ct values obtained from untransfected cells and cells transfected with the ABH1, HFB1 and AQP0 genes when reverse transcriptase was included in the cDNA synthesis reaction. However, nonspecific amplification can't be ruled out as a signal was obtained for AQP0 and ABH1 in both untransfected and transfected cells when reverse transcriptase was omitted. Having said that, for all genes, the height of the bars in Figure 3.2 were the same except for the bar representing transfected cells and RT. This observation was taken as good evidence that the transgenes were being transcribed.

3.2.3. Western Blotting of lysates from transiently transfected CHO S cells

Transfected cells were allowed to grow for 72 hours post transfection, the cells pelleted, washed with ice cold PBS and lysed (Section 2.6.1). The protein content of the lysates

was quantified using a Bradford assay (Section 2.6.2). A sample volume corresponding to 10µg total protein was loaded on a 4-12% denaturing gradient SDS PAGE gel (10µg protein/lane) and run (Section 2.6.3.1) prior to being transferred to a PVDF membrane and detected (Sections 2.6.3.2 and 2.6.3.3). A 6XHis labelled DHFR protein (supplied by Professor Anne Parle Mc Dermot from the School of Biotechnology at DCU) was used as a positive control to demonstrate transfer and blotting had occurred.

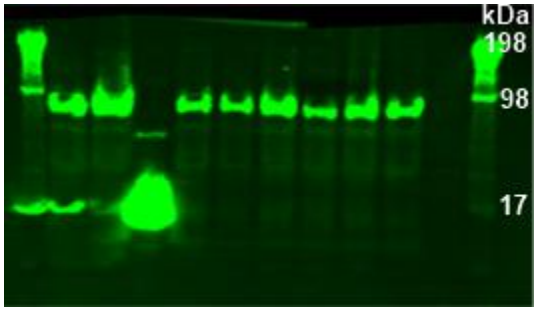
Antibodies against the protein products of ABH1 and HFB1 genes were not commercially available so these genes were synthesised such that the protein product was expressed with the 6XHis label at the C terminus. Hence an anti 6XHis labelled antibody was used to detect these proteins. An antibody against the protein product of AQP0 was commercially available. However, for consistency this gene was also synthesised such that the protein product was expressed with the 6XHis label at the C terminus. Therefore, both the commercial antibody and the anti 6xHis label antibody were used to detect this protein. A lysate of the HT29 cell line was purchased for use as a positive control for the presence of the protein product of AQP0. Figure 3.3a and Figure 3.3b are representative scans of the blots.

If translation of the mRNA transcripts encoding for ABHI, HFB1 and AQP0 had occurred a band at approximately 11.2kDa (ABH1, lanes 5&6), 9.9kDa (HFB1, lanes 7&8) and 28.1kDa (AQP0, lanes 9&10) respectively should have been visible in Figure 3.3a. A band at 28.1kDa (AQP0, lanes 9, 10 and 11) should have also been visible in Figure 3.3b. The reason for the two blots was that the primary antibody used to probe Figure 3.3a was a murine anti his tag antibody while the primary antibody used to probe Figure 3.3b was a rabbit anti AQP0 antibody and consequently different secondary antibodies were required. These secondary antibodies were labelled with a green and red fluorophore (anti-murine and anti-rabbit) respectively.

As can be observed from the blots in Figures 3.3a or 3.3b no bands were observed corresponding to these molecular weights. No bands were observed in the HT29 cell line lysate either. This was not expected as the vendor (Abcam) marketed this cell lysate as a positive control for AQP0. A band was observed for the 6X His labelled DHFR at approximately 21kDa (lane 4). This agrees with the published molecular weight of DHFR indicating that the transfer and blotting steps worked, and the primary and secondary antibody combination used was capable of detecting 6X His labelled proteins.

3.3 A) (His Tagged Proteins)

1 2 3 4 5 6 7 8 9 10 11



Lane 1	**Molecular Weight Ladder**
Lane 2	**Untransfected CHO S Cell Lysate**
Lane 3	**CHO S Cell lysate – GFP**
Lane 4	**His tagged DHFR (Pos. Control)**
Lane 5	**CHO S Cell lysate – ABH1**
Lane 6	**CHO S Cell lysate – ABH1**
Lane 7	**CHO S Cell lysate – HFB1**
Lane 8	**CHO S Cell lysate – HFB1**
Lane 9	**CHO S Cell lysate – AQP0**
Lane 10	**CHO S Cell lysate – AQP0**
Lane 11	**HT29 Cell Lysate (AQP0 Pos. Control)**
Lane 12	**Molecular Weight Ladder**

3.3 B) (AQP0 only)

1 2 3 4 5 6 7 8 9 10 11

Lane 1	**Molecular Weight Ladder**
Lane 2	**Untransfected CHO S Cell Lysate**
Lane 3	**CHO S Cell lysate – GFP**
Lane 4	**His tagged DHFR (Pos. Control)**
Lane 5	**CHO S Cell lysate – ABH1**
Lane 6	**CHO S Cell lysate – ABH1**
Lane 7	**CHO S Cell lysate – HFB1**
Lane 8	**CHO S Cell lysate – HFB1**
Lane 9	**CHO S Cell lysate – AQP0**
Lane 10	**CHO S Cell lysate – AQP0**
Lane 11	**HT29 Cell Lysate (AQP0 Pos. Control)**
Lane 12	**Molecular Weight Ladder**

Figure 3.3: **Western blotting of cell lysates from CHO S cells which had been transfected with the gene(s) of interest. Samples were separated on 4-12% Bis Tris gradient gels. Figure A was probed with a murine anti-6X His tag antibody. Figure B was probed with a rabbit anti-AQP0 antibody. Detection was with anti-mouse and anti-rabbit secondary antibodies which were labelled with green and red fluorophores respectively. Molecular weights on the right hand side for Figures 3.3a and 3.3b were assigned based on the ladders (See Blue® Plus2 prestained) published molecular weight range (Figure 3.3b) and by comparison to Figure 3.3b (Figure 3.3a) .**

The bands seen in all other lanes were attributed to non-specific binding of either the primary or secondary antibody.

3.2.4. Optimisation of the Western Blotting Procedure

Due to the lack of a signal observed with the anti-AQP0 antibody and the HT29 lysate (Figure 3.3b) and the non-specific binding observed in Figure 3.3a some investigation work was performed to try and understand what was causing these issues and to resolve them. The protocol used was shared with Abcam, the vendor of the anti AQP0 antibody and the HT29 cell line. Their response was that they couldn't suggest any optimisation to the procedure followed and they offered a refund. Three experiments to try and explain the non-specific binding observed with the anti 6XHis tag antibody were designed in conjunction with the vendor (Thermo Scientific). The results of these are displayed in Figures 3.4a, 3.4b and 3.4c.

Several aspects of the western blot protocol were changed. The dilution of primary antibody was increased (less primary antibody was used). The anti 6XHis tag antibody was run as a sample on the gel to confirm that the secondary antibody could detect the primary antibody. Additionally, the anti 6XHis tag antibody was spiked into untransfected CHO S cell lysate to determine if other proteins present in the lysate were inhibiting detection of the primary antibody in some way.

In Figure 3.4c the primary antibody was omitted from the detection procedure and only the same secondary antibody as used in Figure 3.4a was used. This blot was run to demonstrate whether the secondary antibody was causing the non-specific binding observed in Figures 3.3a and 3.4a.

In Figure 3.4a the same primary and secondary antibodies as used in Figure 3.3a. It was observed that increasing the primary antibody dilution reduced the non-specific background staining observed in Figure 3.4a compared to Figure 3.3a. The presence of a band at the expected molecular weight (approx. 21kDa) in lane three (which contained the 6XHis tagged DHFR protein) proved that the primary and secondary antibody combination could detect 6XHis tagged proteins. The presence of multiple bands in lanes four and five provided further evidence that the secondary antibody could detect the primary antibody and that CHO S cell lysates don't inhibit this reaction.

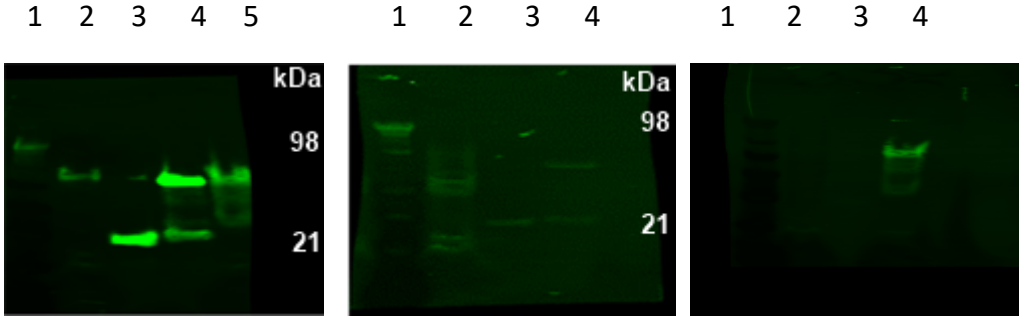
Figures 3.4a and 3.4b used the same samples and secondary antibody but anti His tag antibodies from two different vendors as the primary antibody to rule out the primary antibody as the source of the high background observed in Figure 3.3a. The purpose of

Figure 3.4c was to demonstrate that the secondary antibody could detect the primary antibody and that the presence of CHO host cell proteins in the sample wasn't inhibiting this reaction.

A similar banding pattern was observed in Figures 3.4a and 3.4b with respect to the lanes containing the His tagged DHFR (lane three) and anti His antibody (lane four) in both images. The different banding pattern observed in lane two of Figures 3.4a and 3.4b was attributed to the fact that different primary antibodies were used with these two gels. The poor signal observed in Figure 3.4b compared to Figure 3.4a was attributed to the age of the primary antibody used. This was donated from another group in the university and had been stored in a freezer for several years. The presence of a band at the expected molecular weight (approx. 21kDa) in lane three (which contained the His tagged DHFR protein) in both images proved that the primary and secondary antibody combination used could detect 6XHis tagged proteins. No signal was observed for the positive control (His tagged DHFR, lane 3) in Figure 3.4c as the step where the blot was incubated with the primary antibody was deliberately omitted when analysing this gel. Figure 3.4c provided further evidence that the secondary antibody could detect the primary antibody and that CHO S cell lysates don't inhibit this reaction. This is because the signal detected in lane four was due to the that the primary antibody was included in this sample prior to it being loaded on the gel.

The conclusions drawn from these trouble shooting experiments was that it was the anti 6XHis tag antibody from Thermo Fisher Scientific which was responsible for the non-specific binding and that the presence of other proteins in the CHO S lysate didn't interfere with detection of the primary antibody by the secondary antibody.

3.4 A) **3.4 B)** **3.4 C)**



Lane 1	**Molecular Weight Markers.**	**Molecular Weight Markers**	**Molecular Weight Markers**
Lane 2	**Untransfected CHO S Cell Lysate.**	**Untransfected CHO S Cell Lysate**	**Untransfected CHO S Cell Lysate**
Lane 3	**His tagged DHFR protein (positive control).**	**His tagged DHFR protein (positive control)**	**His tagged DHFR protein (positive control)**
Lane 4	**Anti His antibody (Thermo product MA1-21315) run as a sample on the gel.**	**Anti His antibody (rabbit anti His (Abcam ab1187) run as a sample on the gel.**	**Untransfected CHO S Cell Lysate spiked with Thermo product MA1-21315**
Lane 5	**Untransfected CHO S Cell Lysate spiked with Thermo product MA1-21315.**	**Empty**	**Empty**

Figure 3.4: Western blot method optimisation. Samples were separated on 4-12% Bis Tris gradient gels. All blots were probed with a murine anti-6X His tag antibody. Detection was with a secondary anti-mouse antibody labelled with a green fluorophore. Figures 3.4a and 3.4b differ in the supplier of the primary (anti 6X His tag) antibody. Figure 3.4c spiked the anti 6X tag antibody into CHO S supernatant and omitted it from the WB procedure. The kDa values on the right hand side of Figures 3.4a and 3.4b correspond to the single band visible in lane 1 and lane 3. As no bands are visible in lanes 1 or 3 molecular weights have not been added to Figure 3.4c.

Given the fact that the primary antibody used in this section was causing non-specific binding in CHO S cells and as expression of the three proteins which had been transfected into CHO S cells couldn't be demonstrated by western blotting it was decided to change the host cell line to CHO K1 and the transfection reagent to *Trans*IT-X2 (Mirus Bio) on the basis that these were the host cell line and transfection reagent in routine use in our lab. Therefore, parental CHO K1 suspension derived cells were obtained from internal cell banks and were revived as described (Section 2.2.3).

3.2.5. Transgene Expression in Transiently Transfected CHO K1 cells

One mL of CHO K1 cells @1X10^6 cells/mL were transfected with 500ng total DNA using TransIT-X2-Dynamic Delivery System (Mirus Bio) as per Section 2.4.1. Transient transfections were performed in twenty-four well plates and allowed to grow for 24 hours before 500µL of cell suspension was removed for RNA isolation. The remaining cell suspension was refed with fresh pre warmed media and returned to the incubator for a further 72 hours to allow for protein expression.

DNase treated RNA, isolated from transiently transfected CHO K1 cells, was used as the template in a cDNA synthesis reaction with random primers and reverse transcriptase. This was subsequently used as the template in a PCR reaction with primers specific for GAPDH, ABH1, HFB1 and AQP0 respectively (Figure 3.5). A cDNA synthesis reaction excluding the enzyme reverse transcriptase was used as a negative control for each reaction.

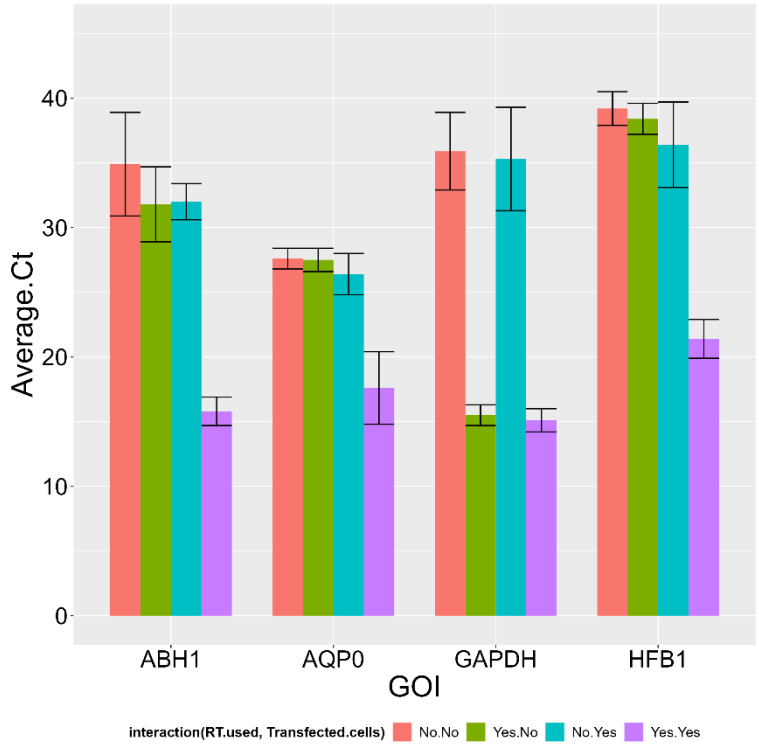


Figure 3.5: **Average Ct values obtained for target genes following RNA isolation from transfected and untransfected CHO K1 cells and subsequent cDNA synthesis (including and excluding reverse transcriptase). The graphs are the average of three biological replicates and each biological replicate was itself the average of three technical replicates.**

A similar profile to that observed in Figure 3.2 was observed for ABH1, HFB1 and AQP0 following transfection into CHO K1 cells. Template synthesised without reverse

transcriptase should have had a Ct value ≥ 35 but this was only observed for GAPDH and HFB1. The difference between the Ct values obtained for ABH1 when reverse transcriptase was included (15.8) or omitted (32.0) from the cDNA synthesis reaction was 16.2 or 4.9 $\log_{10}$. Similarly, the difference between the Ct values obtained for APQ0 when reverse transcriptase was included (17.6) or omitted (26.4) from the cDNA synthesis reaction was 8.8 or 2.7 $\log_{10}$.

The Ct values obtained when cDNA from untransfected CHO K1 cells was used as the template with the AQP0 primer set gave similar values to those obtained with transfected cells without reverse transcriptase. This matched the pattern observed for the transfected and untransfected CHO S cells. In both cell lines all genes were detected several cycles earlier than the control, providing good evidence of successful transcription.

To provide confidence that the data from the RT-qPCR experiments represented amplification of the target sequence and was not solely due nonspecific amplification, the $\log_2$ of the fold change, $\text{Log}_2(-\Delta\Delta\text{Ct})$, for each gene was calculated using untransfected cells as the control and transfected cells as the test. The results of this calculation are presented in Figure 3.6. The fold change values ranged from 2.7 (AQP0) to 4.0 (HFB1). A positive fold change is proof that a greater amount of template was present in the transfected cells compared to the untransfected cells.

Additionally, for the AQP0 gene only, a one tail, one sample Wilcoxon test was performed to examine if the difference between the average of the three Ct values from untransfected and transfected cells, having included reverse transcriptase in the cDNA synthesis step, was significant. This test was chosen due to the small sample size and as it's the non-parametric equivalent of the one tail, one sample student t test. The result of the Wilcoxon test was $p = 0.0004$ (CHO S) and $p = 0.005$ (CHO K1). As this value was < 0.01 it was concluded that the signal observed in the RT-PCR experiments for AQP0 cDNA synthesised using reverse transcriptase from transfected cells was due to the presence of the transfected gene and not nonspecific amplification in both cell lines. It seemed reasonable to conclude that this increased amount of template in the transfected cells was due to transcription of the transfected gene as all other variables are constant between transfected and untransfected cells. If there was contamination in the RT-qPCR mixture surely it would have affected the Ct values obtained for the

transfected cells equally since the only difference between these two preparations was the inclusion or exclusion of reverse transcriptase for the cDNA synthesis step. This was not the case in either CHO S or CHO K1 cell lines.

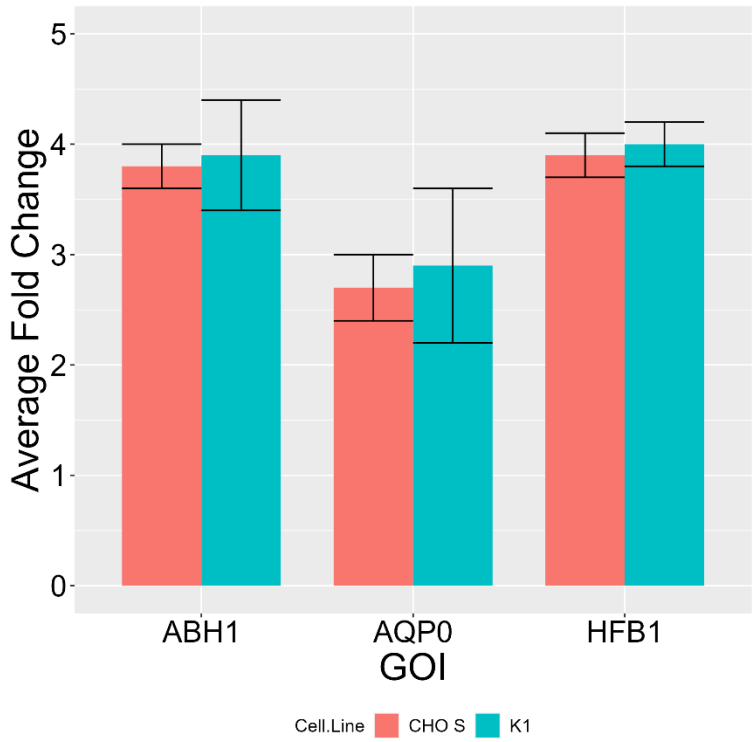


Figure 3.6: Average (3 biological replicates) fold change in expression of the GOI in transfected cells versus untransfected cells obtained by calculating $\log_2(-\Delta\Delta Ct)$ following normalisation against GAPDH.

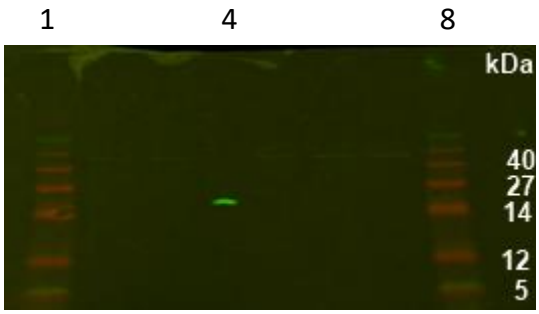
3.2.6. Western Blot Analysis of transfected CHO K1 Cells and supernatants

Having concluded that the transfected genes were being transcribed the next step was to see if the translated protein could be detected in cell lysates. Transfected cells were allowed to grow for 72 hours post transfection, the cells pelleted, washed with ice cold PBS and analysed as per section 2.6.3.

Figures 3.7a and 3.7b are representative images of western blots of whole cell lysates for all three transfected genes from one round of transfection and AQP0 cell culture supernatant samples from all three rounds of transfection. The blots showing the whole cell lysates for the other rounds of transfections and the cell culture supernatant samples for ABH1 and HFB1 showed a similar pattern.

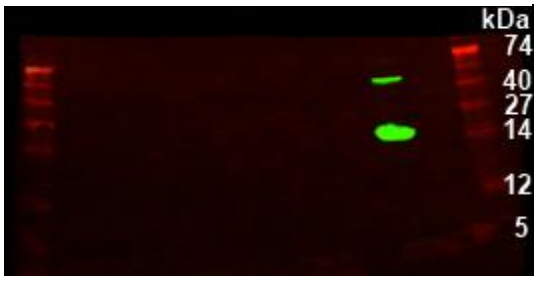
If translation of the mRNA transcripts encoding for ABHI, HFB1 and AQP0 had occurred a band at approximately 11.2kDa (ABH1, lane 5, figure 3.7a), 9.9kDa (HFB1, lane 6, figure 3.7a) and 28.1kDa (AQP0, lane 9, figure 3.7a and lanes 2, 4 and 6, figure 3.7b) respectively should have been visible.

3.7 A)



Lane 1	**Molecular Weight Ladder**
Lane 2	**Untransfected CHO K1 Cell Lysate (Neg Control)**
Lane 3	**CHO K1 Cell Lysate – Empty Vector**
Lane 4	**His tagged DHFR (Pos. Control)**
Lane 5	**CHO K1 Cell lysate – ABH1**
Lane 6	**CHO K1 Cell lysate – HFB1**
Lane 7	**CHO K1 Cell lysate – AQP0**
Lane 8	**Molecular Weight Ladder**

3.7 B)



Lane 1	**Molecular Weight Ladder**
Lane 2	**CHO K1 Supernatant – AQP0**
Lane 3	**Blank**
Lane 4	**CHO K1 Supernatant – AQP0**
Lane 5	**Blank**
Lane 6	**CHO K1 Supernatant – AQP0**
Lane 7	**Blank**
Lane 8	**Neat Media**
Lane 9	**Blank**
Lane 10	**His tagged DHFR (Pos. Control)**
Lane 11	**Blank**
Lane 12	**Molecular Weight Ladder**

Figure 3.7: Representative Western blots of cell lysates from CHO K1 cells (a) and supernatants (b) which had been transfected with the gene(s) of interest. Samples were separated on a 12% Run Blue™ Bris Tris Gels. Blots were probed with a murine anti-6X His tag antibody. Detection was with a secondary anti-mouse antibody labelled with a green fluorophore. Molecular weights on the right hand side for Figures 3.7a and 3.7b were assigned based on the molecular weight ladder's (Run Blue™ prestained) published range.

A band was observed for the 6X His labelled DHFR at approximately 21kDa (lane 4). This agrees with the published molecular weight of DHFR indicating that the transfer and

blotting steps worked, and the primary and secondary antibody combination used was capable of detecting 6X His labelled protein.

3.2.7. Expression of a AQP0 -GFP Fusion Protein

It could be argued that difficulties in expressing the hydrophobin proteins (if not he mRNAs) could have been expected. These are fungal proteins, rich in hydrophobic and cysteine residues and as such are ideal candidates for membership of the category of "difficult to express" proteins. However, there's no reason why CHO S and K1 cells shouldn't be able to express the AQP0 protein. This is a mammalian protein and its expression in murine cells has been reported (Kumari and Varadaraj, 2009).

In collaboration with a fellow PhD student (Ricardo Valdés Bango Curell) one final attempt was made to see could expression of AQP0 be detected in CHO cells. An AQP0-GFP fusion protein was generated by fusing GFP to the C terminus of the AQP0 gene so that translation/expression of the GFP protein could only occur if the AQP0 protein was also translated. This vector was then transfected into CHO K1 cells. Briefly CHO K1 cells were used to seed Lab-Tek Chambered #1.0 Borosilicate Cover Glass System (Part No 155411) @ $6.25X10^4$ cells/mL in 400µL CHO S SFM II media supplemented with 10% FBS. Cells were grown for 24 hours @ 37°C, 5% CO_2 in a static humidified chamber prior to transfection. On the day of transfection, the media was removed and replaced with 400µL CHO S SFM II media without FBS. The cells were transfected with 50ng of the AQP0-GFP construct and returned to the incubator for a further 24 hours before imaging them. A plasmid expressing only GFP (N44) was transfected as a control.

Cells were imaged at 60X under oil immersion with the appropriate filters for GFP fluorescence. The results are presented in Figures 3.8a and 3.8b. The AQP0-GFP fusion was constructed such that transcription and translation of the AQP0 sequence had to occur for a functional GFP protein to be expressed. Figure 3.8b clearly demonstrates that CHO K1 cells can express fully functional human AQP0 and furthermore the inclusion of AQP0 drives the localisation of GFP at the cell membrane in comparison to Figure 3.8a whereas GFP is dispersed throughout the cell's cytoplasm. This provides further evidence that the lack of signal detected in the western blots (Figures 3.3 and 3.4 and 3.7) was due to technical issues rather than a lack of protein expression. The work described in this section was performed towards the end of the PhD and a decision had already been made by the project sponsor to halt work on this part of the thesis and

progress the work described in chapters 4-7 inclusive. This decision was based on the fact that the sponsor had defined success as substantial flocculation (>50% of cells clumped) and this was not observed during any of the transient transfections performed during the experiments described in this chapter.

3.8 A) **3.8 B)**

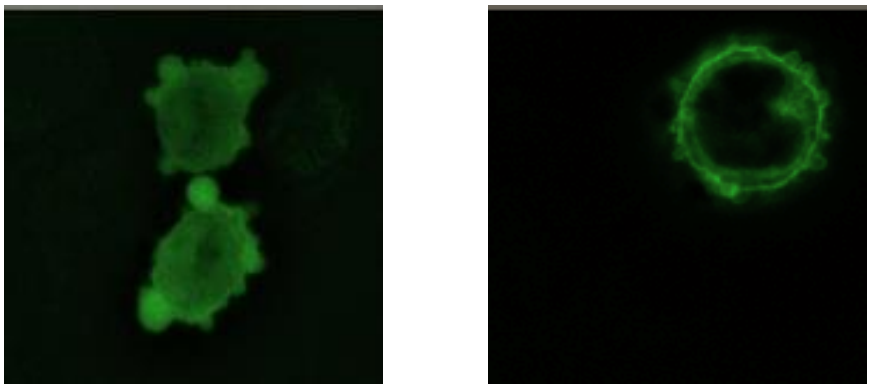


Figure 3.8: Fast Z stacked images of CHO K1 cells transfected with GFP only (a) and an AQP0-GFP fusion protein (b). Imaged at 60X, with filters for GFP. GFP was excited at (peak) 470nm and emitted at 509 nm.

3.2.8. Cloning and attempted expression of members of the Flo family of proteins

This project started with five candidate genes which were thought to have the potential to flocculate cells. The final two candidates, Flo5 and Flo11 are both members of the Flo family of genes from *S. cerevisiae*. The ability of this family of genes to drive the flocculation of *S. cerevisiae* cells at the end of fermentation in the brewing industry is well documented. Unfortunately, the advice of several gene synthesis companies was that due to the presence of repeating nucleotide sequences in both genes there was an exceptionally low chance of successfully synthesising/expressing them. Given that Bidard et al. (1994) reported a linear correlation between the number of repeats and the degree of flocculation it appeared as if these genes wouldn't be suitable candidates. However, Veelders et al. (2010) demonstrated that only the first 271 amino acids of the Flo5 gene were necessary for flocculation. Consequently, this sequence was synthesised and used to transfect CHO K1 cells. The results are displayed in Figure 3.9.

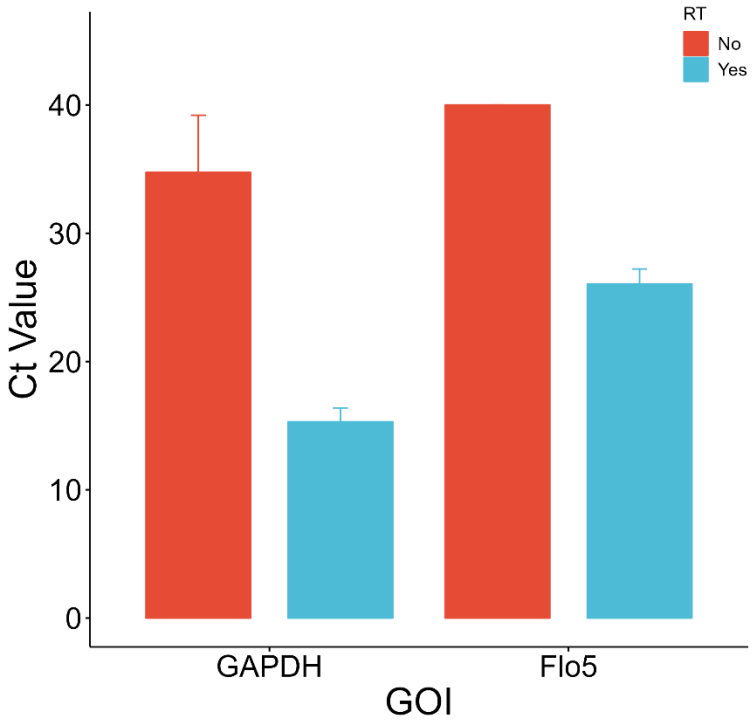


Figure 3.9: **Average Ct values obtained for cloned genes following RNA isolation from transfected CHO K1 cells and subsequent cDNA synthesis (average Ct value is based on three biological replicates).**

The PCR analysis (Figure 3.9) clearly showed that the Flo5 sequence was transcribed by the cells albeit at a modest level (Ct = 26.1) compared to GAPDH (Ct = 17.5). No further analysis was performed on these cells as explained in the paragraph below.

Given the challenges experienced with the earlier three targets, a decision was made to discontinue this approach. This decision was also informed by the project sponsor's desire to see significant flocculation being caused by these targets. As there was no apparent cell aggregation evident in these transient transfections, despite evidence of mRNA expression, this threshold seemed unlikely to be met. Therefore, alternative approaches were considered. These approaches are reported in chapters 4-7.

3.3. Discussion

3.3.1. Cloning and Expression of ABH1, HFB1 and AQP0 Genes in CHO Cell Lines.

The proteins chosen as potential novel biological flocculating agents in this project were diverse in their origin. Hydrophobins and the Flo family of proteins possess characteristics common to many difficult to express proteins (e.g., presence of multiple cysteine residues or contain multiple tandem repeats) and their exogenous expression in mammalian cells has not previously been reported either because: it's never been attempted or, it has been attempted but has failed, and therefore never been reported. Despite this hydrophobins have been successfully expressed in yeast (Niu et al., 2012) and insect (Lahtinen et al, 2008) based expression systems.

Three genes (ABH1, HFB1 and AQP0) were cloned in two CHO cell lines (CHO S and CHO K1). A fourth gene (Flo5) was cloned in CHO K1 only. The purpose of performing these RT-qPCR experiments was to simply confirm that the transfected genes were being transcribed, not to quantify the amount of transcription occurring. Therefore, it was deemed acceptable to prove that there was a difference in Ct values obtained from transfected versus untransfected cells having normalised by reference to an endogenous control gene. GAPDH is a commonly accepted endogenous gene which fulfilled that function in this study. Across all 3 rounds of transfections, for both CHO S and CHO K1 cells, GAPDH had an average Ct value of 15.4 and 15.1 respectively. The %RSD for the average Ct value was 3.9% and 6.1% respectively. Having performed this normalisation step the fold change for all genes between untransfected and transfected cells was >2.0 in both cell lines. A fold change of two is generally considered as the minimum acceptable fold change in RT-qPCR experiments. The fact that the fold change for both the ABH1 and AQP0 genes was > 2.0 was taken as sufficient evidence that the RT-qPCR results showed that the transfected gene was being transcribed as opposed to being caused solely by nonspecific amplification (Hellemans and Vandesompele, 2011).

RNA from transfected/untransfected cells could also have been sent away for RNA seq analysis which would have provided absolute confirmation that the cloned GOI were being transcribed in the transfected cells.

3.3.2. Exogenous Protein Expression and its Detection in CHO Cell Lines

Western blot optimisation experiments proved that the primary and secondary antibody combination used could detect 6XHis tagged proteins. However, no 6XHis tagged proteins were detected in cell lysates of transfected CHO S cell lines or cell lysates and supernatants of CHO K1 cell lines. There could be several reasons for this.

1. The poly 6XHis tag wasn't expressed on the surface of the folded protein and consequently wasn't available for the primary antibody to bind to. Fusion of the poly 6XHis tag to the C terminus was chosen so that if a signal was detected by western blot, we would have confidence that full length proteins were being detected.

2. In their parental species, both ABH1 and HFB1 are localised at the growing hyphal tip. Similarly Figure 3.9 proved that AQP0 was expressed at the cellular membrane of transfected CHO K1 cells. The cell lysates used in the western blotting experiments were whole cell lysates where no enrichment step of membrane bound/associated proteins occurred. Although some membrane proteins can be detected in whole cell lysates an enrichment step is often performed due to the hydrophobic nature of many membrane proteins. The only reason these western blots weren't repeated using membrane enriched lysates was that a decision had been taken to focus on other areas given the lack of a flocculating phenotype in transfected cells.

3. Hydrophobins are characterised by the presence of eight conserved cysteine residues which form four disulphide bonds. Prior to self-assembly into amphipathic membranes, at an air water interface, Hydrophobins have been shown to be soluble at concentrations ≤ 100mg/mL (Kisko et al., 2008). Additionally, there's evidence (de Vocht et al., 2000) that it's the disulphide bridges formed by these cysteine residues that mediates hydrophobin solubility in water by preventing self-assembly into insoluble amphipathic membranes. The cytosol of mammalian cells is recognised as a reductive environment and therefore it's possible that these disulphide bridges get reduced thus negatively impacting the ability of CHO cells to express soluble monomeric hydrophobins. There's some evidence for this theory when using *E. coli* as the expression

system. Cui et al. (2020) successfully expressed the class I hydrophobin EAS (rodlet protein) from *Neurospora crassa* and class II hydrophobin HFBII from *T. reesei* in the periplasmic space in *E. coli* cells. The periplasmic space is known to be an oxidative environment which supports the formation of disulphide bonds. Similarly, Siddiquee et al. (2020) expressed multiple, correctly folded, soluble class 1 and class 2 hydrophobins using a cell free *E. coli* expression system, thus avoiding the issue of hydrophobins being expressed as insoluble inclusion bodies when using *E. coli* as the host (Morris et al., 2013).

4. CHO cells lacked the cellular machinery (chaperones) to process and correctly fold fungal proteins. Accumulation of incorrectly folded proteins in the endoplasmic reticulum of CHO cells is known to result in the activation of the unfolded protein response pathway (UPR), reviewed by Schröder (2006). This is a system by which cells get rid of misfolded proteins. It would be interesting to see whether the transfected cells displayed an elevated UPR by quantifying the expression of genes such as ATF6, ERAD, etc. (see Sicari et al., 2020) for a review of approaches for assessing ER stress in mammalian cells).

5. The mRNA signal was unstable and hence translation was compromised. It's known that the individual codons which make up a gene's sequence impact on mRNA stability. Previous work on codon optimisation had suggested that certain codon sequences were preferred for individual amino acids. For example, alanine is encoded by the nucleotide sequence GCC, GCT, GCA and GCG. The difference in the usage frequency of GCC is five times greater than that of GCG in CHO cells (taken from the Genescript.com website). Hence codon optimisation of gene sequences prior to cloning has become a standard approach in the biologics industry. However, Forrest et al. (2020) linked individual amino acids and mRNA stability in CHO cells. They demonstrated that it's the amino acid rather than the codon which determines mRNA stability. For example, the amino acid alanine was associated with an overall stabilising effect while cysteine tended to have a destabilising effect. This is interesting in the context of this work as hydrophobins possess a characteristic pattern of eight cysteine residues and make up approximately 8% of the amino acids in a hydrophobin.

6. Translation was not being initiated for some other reason e.g., the concept of "difficult to express" proteins, reviewed by Alves and Dobrowsky (2017).

All these points could explain why the proteins encoded by the ABH1 and HFB1 genes weren't detected when using an anti 6XHis tagged antibody. The results of the AQP0-GFP fusion protein experiment points towards point two being the reason why no signal was detected when cell lysates of CHO cells transfected with the sequence for AQP0 were analysed by western blotting.

3.3.3. Alternative Strategies to express these proteins in CHO cells

1. An option that was considered at the start of the project but rejected was expressing the genes investigated in this chapter as fusion proteins. There are reports in the literature that hard to express proteins can be expressed if they are fused in frame with the C terminus of another protein e.g., SUMO or maltose binding protein (MBP). Despite evidence that SUMO works in partnership with ubiquitin to target proteins to the proteasome, Tatham et al. (2011), Peroutka et al. (2008) demonstrated that conjugation of toxic phospholipase A2 from cobra venom increased the expression levels of functionally active phospholipase A2 secreted into the media using a HEK-293T cell line. Similarly, Reuten et al. (2016) demonstrated that MBP performed best in a comparative study of several expression tags using a HEK 293 cell line. They observed that cells transfected with MBP fusion proteins had higher cell viabilities post transfection compared to controls. Their data showed that the presence of the MBP tag was sufficient to traffic the fusion protein through the ER and Golgi apparatus in the absence of a signal peptide. They hypothesised that fusion to MBP resulted in accelerated protein folding rates or quicker mRNA turnover during translation of the fusion protein.

This approach was rejected for the ABH1 and HFB1 genes. As these genes encode proteins <100 aa in size it was felt that expressing them as fusion proteins could potentially mask their unique properties which this project hoped to exploit. Also, this approach was considered unlikely to be necessary since the AQP0 gene encoded a human protein. Ironically, it was by expressing the AQP0 gene as a

fusion protein with GFP that expression of the AQP protein was demonstrated and its subcellular location determined.

2. Codon optimisation, a recognised tool for increasing the expression of hard to express proteins in CHO cell lines (reviewed by Mauro and Chappell, 2014), was also considered but rejected at the start of the project. The reason behind this is that it was felt that using less favoured codons might facilitate expression of the hydrophobin proteins, ABH1 and HFB1, as the cellular machinery would be slowed down and the cell would have time to adapt to producing these unfamiliar proteins. There's evidence that this is a valid approach. Siddiquee et al (2020) demonstrated that the concentration of plasmid DNA used during cell free expression of hydrophobins in *E. coli* impacted on hydrophobin expression levels. Interestingly they reported that lower DNA concentrations resulted in increased hydrophobin expression and attributed this to "lower DNA concentrations allowing for a slower rate of protein production which gives time for the four disulphide bonds to form correctly and for the hydrophobins to fold, thereby avoiding protein aggregation and precipitation." In hindsight the sequence for the AQP0 gene should have been codon optimised for expression in CHO cells prior to cloning as codon optimisation is standard procedure when expressing proteins in CHO cells.
3. The approach taken by Laux et al. (2013) in which they combined gene expression profiling with genetic engineering to increase the expression of their difficult to express protein by a factor of approximately 6-fold from 50mg/L to > 300mg/L. In this instance gene expression profiling identified downregulation of mitochondrial genes, along with upregulation of a gene (gene A) which potentially triggered this downregulation. They used zinc finger nucleases to create a knockout cell line which lacked gene A. The growth profile of this knockout cell line was comparable to the wild type cell line with increased expression of the desired protein.
4. A standard off the shelf expression vector (pcDNA3.1) was used in this project. However, there is a growing body of experimental evidence that a "one size fits all" approach may not be appropriate in all cases. Therefore, to maximise expression of heterologous genes in CHO cells, attention needs to

be paid to the specific demands expression of a given heterologous protein places on the host cell as well as vector design and the position(s) in the host genome where the cloned gene inserts. The latter can have a significant effect on the activity of the cloned genes promoter. Various screening approaches for modelling the effects of different approaches on difficult to express proteins have been described in the literature recently, (Pristosvšek et al., 2019 and Cartwright et al., 2020) which reinforce this point.

5. Knight et al. (2021) investigated whether adding inhibitors of the proteasome to cell culture media increased protein expression of a standard chimeric IgG_4 (cB72.3) antibody and a known difficult to express fusion protein. In both cases inclusion of proteasome inhibitors in the culture media increased the expression of the endogenous protein.
6. Stable cell lines could have been created as opposed to performing transient transfections. Although it's standard practice to use transient transfections in proof-of-concept studies it might have paid off to create stable cells lines from the start of this project. The issue with transient transfections was that unless a selective pressure was exerted, expression of the transfected gene decreases over time as the plasmid is lost. Creation of stable cell lines would have had two advantages:
    - The GOI would have been integrated into the CHO cell genome. This would have allowed fed batch cultures to be run over extended (10 days or more) periods of time. This would have allowed a greater length of time for translation of the GOI to occur and hence higher levels of protein to be produced. This could have provided a definitive answer to the question was the transfected protein being expressed? It's also possible that a critical threshold of transfected protein was required to drive flocculation and this threshold was not exceeded in the transient transfections reported in this study.
    - Transfection systems which insert the GOI into a transcriptionally active "hot spot" e.g., by using site specific recombinases (Phan et al., 2017) or TALEN and CRISPR/Cas9 (Kawabe et al., 2018) could have been exploited. The approach described by Phan lead to a greater than 19-fold increase in expression of human bone

morphogenic protein 2 (hBMP2) in CHO K1 cells compared to controls. While hBMP2 is not a difficult to express protein it's a useful example of the types of gains in expression that can be achieved using site specific integration approaches.

3.4. Conclusion

It's important to note that the work undertaken in this chapter was always recognised by the industrial sponsor as having a high risk of failure but that it represented a novel approach to solving a pressing issue facing the biologics industry. This made it worth investigating particularly as modern flocculants such as pDADMAC can't be used in legacy multi product stainless-steel plants because they adhere to the bioreactor surface and are extremely difficult to remove thus increasing the risk of cross contamination between batches.

CHAPTER 4

Flocculation of CHO K1 cells with the inorganic salt of various cations and an assessment of the impact on the product quality of a model antibody

4.1. Introduction

In chapter 3 a cell engineering approach to cell flocculation was presented. The cell engineering approach was unsuccessful as expression of the cloned proteins could not be detected by western blot and flocculation of the transfected cells compared to untransfected controls was not observed. As an alternative approach towards developing a "one approach fits all scales" primary recovery method the use of inorganic divalent (calcium, magnesium and manganese) and trivalent (iron(III)) cation salts as flocculating agents for use with CHO K1 cells was investigated as they are known to aid flocculation of mammalian cells (Romero et al., 2015 and Keszey and Sütö, 2020) Other examples include the role calcium (Taylor and Orton, 1973, 1975; Stratford, 1989), magnesium and manganese (Stewart and Goring, 1976 and Miki et al., 1982) play in the flocculation of yeast during brewing. Similarly, magnesium chloride has been shown to play a key role during cellular adhesion to surfaces (Takeichi and Okada, 1972). The use of inorganic cations such as these would have an advantage over synthetic polymers such as pDADMAC as there would be no regulatory requirement to demonstrate clearance by the purification process. Additionally, while flocculants such as pDADMAC are not the most expensive consumable used during the manufacture of recombinant proteins, the use of inorganic cations does have a cost advantage over synthetic polymers. Based on the paper by McNerney in 2013 approximately 120L of 10% pDADMAC would be required to flocculate a 20,000L bioreactor. This would cost between three and four million euro based on the list price at time of writing. The total cost of goods is an industrywide approach to measuring how expensive it is to produce a gram of recombinant protein. Given that the biotech industry globally has a goal of driving the total cost of goods to < US$10/g recombinant protein the replacement of synthetic polymers such as pDADMAC with inorganic cations can help meet this goal.

It's important to make an observation about magnesium chloride. Magnesium chloride has been used as a flocculant in both wastewater treatment plants (Semerjian and Ayoub, 2003) and the food industry (Li Tay et al., 2006). Westoby (2011) demonstrated that magnesium chloride caused a reduction in the turbidity of CHO cell cultures at pH values between pH 5 and pH 6 compared to pH 7 and its use as a chemical flocculant in combination with the commercial flocculant pDADMAC was investigated (Leng et al.,

2015). Despite this it wasn't used in this study. The reason behind this omission was based on safety grounds. To minimise the volumes of flocculant required 5M was chosen as the stock concentration (at a 100mM final concentration, 400L of a 5M flocculant solution would be required to flocculate a 20,000L reactor). Dissolution of magnesium chloride in water is a highly exothermic reaction and it's not possible to safely make up a 5M stock. Obviously, less concentrated stock solutions could be used but there is a trade off between cation concentration and the volume of the bolus shot required to induce flocculation (halving the stock concentration would double the bolus volume).

In this chapter the ability of calcium chloride, manganese chloride and ferric chloride to promote flocculation of CHO K1 cells was examined at a range of cation concentrations and at two different total cell concentrations. The cell concentrations chosen were $8X10^6$ cells/mL and $2X10^7$ cells/mL. $8X10^6$ cells/mL represented the typical cell concentrations encountered in industrial processes while $2X10^7$ cells/mL was chosen as being representative of an industrial process with higher cell concentrations. Given that acidification of the bioreactor broth prior to commencing harvest operations is a known strategy in the biologics industry the impact of culture pH on flocculation was also considered.

The experimental approach that was followed was broken down as follows:

- Initial screening experiments were performed to investigate the effect of cation concentration on the number of cells remaining in solution and cell culture viability at two time points (15- and 60-minutes post cation addition). These experiments were performed at two different total cell culture concentrations and culture pH values.
- The results of this initial screening informed the concentrations used in a second round of experiments which investigated the effect of cation concentration on sedimentation rate.
- Finally, the impact of this approach on a recombinant protein was mimicked by adding a model antibody (human IgG purchased from Abcam) to non-producing CHO cells, sedimentating the cells, purifying the IgG and performing titre (ELISA), purity (SDS-PAGE) and size (SE-HPLC) analysis. Analysis of charge variants (oxidation etc) was not performed as the facilities to perform isoelectric focusing

experiments, which is the standard method used in industry, were not available at DCU.

4.2. Results

An initial set of range finding experiments were conducted to identify the effect of cation concentration on the flocculation of CHO K1 cells in different media. The medium used to grow the CHO K1 cells was Balan CD, and the two variations were “spent media” and “fresh media.” Cultures were sampled at 15 and 60 minutes after the addition of cation. These timepoints were chosen as being representative of the timeframe that would be acceptable in the biologics industry. Sixty minutes was set as the maximum time for flocculation based on reports of antibody dissociation during primary recovery (Hutterer et al., 2013). Initially, cations were evaluated, in both spent and fresh media, at three concentrations and two cell densities, $2X10^7$ cells/mL and $8X10^6$ cells/mL which were deemed to be representative of the range of cell concentrations experienced in industry. The cation concentrations (5-500mM) examined were chosen as covering what was practical to use in industry at large scale (>1000L) while including an extreme outlier (500mM) in the initial range finding experiment to ensure flocculation was observed. However, no data was obtained for the 500mM concentration in either spent or fresh media for any of the tested cations. This was because the amount of clumping at this concentration was so great the Cedex XS system couldn’t analyse them. Following review of the results the concentrations for each individual cation were refined and further experiments used only spent media and $2X10^7$ cells/mL. This decision was based on the finding that, contrary to expectation, the degree of flocculation was greater in fresh media compared to spent media and therefore spent media and $2X10^7$ cells/mL represented a worst-case scenario. Additionally, a media change prior to commencing primary recovery is not a practical option in commercial environments using fed batch STRs. A list of these initial experiments is presented in Table 4.1. The amount of variability in the data at 15 minutes and whether there was a significant effect compared to the negative control was assessed by the Kruskal Wallis test. This test was chosen due to the small sample size and as it’s the non-parametric equivalent to a one-way AVOVA with $p \leq 0.05$. Pairwise comparisons between the individual cation concentrations and the negative control were assessed using the Wilcoxon test (a non-parametric equivalent to Tukeys HSD) at a $p \leq 0.05$ using a Benjamini-Hochberg adjusted p value.

Media Type	Cation ID	Cation Conc. (mM)	Time (min)	$2X10^7$ cells/mL	$8X10^6$ cells/mL
Spent	Calcium ($CaCl_2$)	500	15 and 60	Y	Y
		200		Y	N
		100		Y	N
		50		Y	Y
		5		Y	Y
Spent	Manganese ($MnCl_2$)	500	15 and 60	Y	Y
		50		Y	Y
		25		Y	N
		10		Y	N
		5		Y	Y
Spent	Iron ($FeCl_3$)	500	15 and 60	Y	Y
		50		Y	Y
		25		Y	N
		10		Y	N
		5		Y	Y
Fresh	Calcium ($CaCl_2$)	500	15 and 60	Y	Y
		50		Y	Y
		5		Y	Y
Fresh	Manganese ($MnCl_2$)	500	15 and 60	Y	Y
		50		Y	Y
		5		Y	Y
Fresh	Iron ($FeCl_3$)	500	15 and 60	Y	Y
		50		Y	Y
		5		Y	Y

Table 4.1: **Summary of the different cation and concentrations used in the flocculation experiments. "Y" and "N" refer to whether this condition was evaluated (Y) or not (N).**

The experiments described in this chapter weren't designed to provide evidence for a flocculation strategy based on changing the media prior to harvest. The idea behind "fresh media" was that because these experiments were performed on batch cultures with no pH control, metabolic waste products would accumulate in the media during

growth, and this could impact on pH. Pelleting the cells and resuspending in fresh media prior to the addition of cation, allowed the effect of pH on flocculation to be mimicked.

4.2.1. Effect of Calcium Chloride Concentration on Total Cell Number in Spent and Fresh Media (2X10^7 cells/mL)

The overall amount of variability in the data at 15 minutes and whether a significant effect was observed compared to the negative control was assessed by the Kruskal Wallis test ($p = 0.01$, spent media and $p = 0.03$, fresh media). Even though the number of total cells/mL remaining in solution decreased with increasing cation concentration (Figures 4.1, 4.2a and 4.2b) none of the tested concentrations resulted in a significant effect compared to the negative control in either fresh or spent media when pairwise comparisons were performed. The data does not provide evidence to support the hypothesis that 50mM calcium chloride in spent medium will significantly promote cellular flocculation. Therefore, it was decided to investigate the effect of 200mM and 100mM calcium chloride on total cell number/mL remaining in solution after incubation with cation in spent media only. No data was plotted for the 200mM concentration. This was because the amount of clumping in this sample was so great the Cedex XS system couldn't analyse it. The 100mM concentration had no significant effect compared to the negative control in spent media (Figures 4.1, 4.2a and 4.2b**)**. For fresh media, the same trend of a decrease in the number of total cells/mL remaining in solution with increasing cation concentration was observed.

Regarding cell viability, the use of calcium chloride as a flocculating agent with 2X10^7 cells/mL only impacted on cell viability at 100mM in spent media and 50mM in fresh media (Figure 4.1)

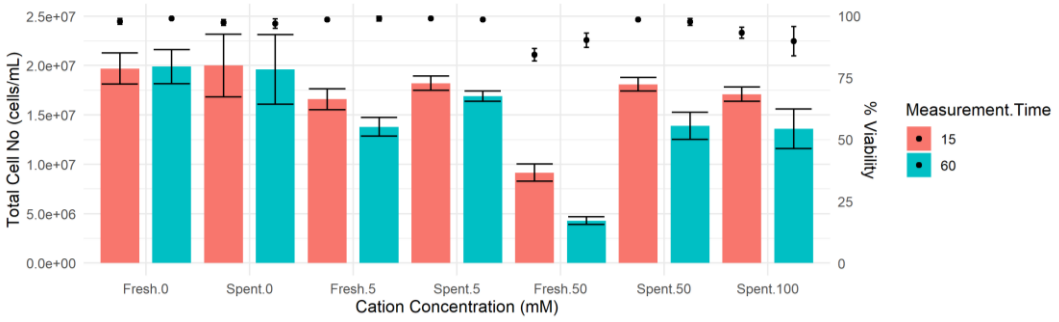


Figure 4.1: **Total cell number remaining in solution (bars) and % viability (dots) after 15 and 60-minutes incubation of CHO K1 cells with calcium chloride. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in either the same media (spent) or new media (fresh) at 2X10^7 cells/mL. A 10X solution of calcium chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm prior to being sampled. The bars/dots representing the spent zero point are the mean of six biological replicates, all others are the mean of three biological replicates. The error bars represent the standard deviation of those replicates.**

Figure 4.2a

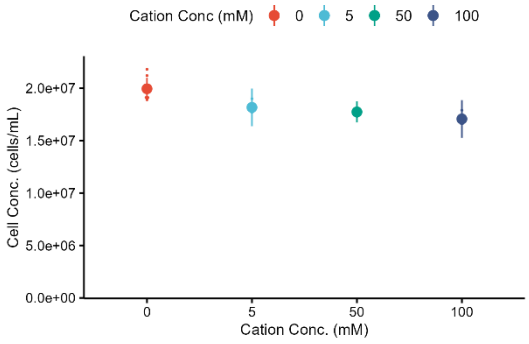


Figure 4.2b

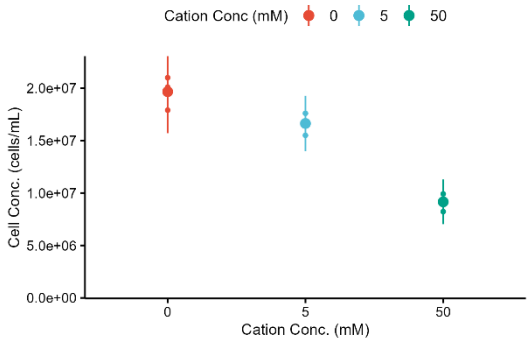


Figure 4.2: **Total cell number remaining in solution measured after 15 minutes incubation with calcium chloride. Each dot is an individual biological replicate. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent, Figure 4.2a) or fresh media (fresh, Figure 4.2b) at 2X10^7 cells/mL. A 10X solution of calcium chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm. The bars on the plot represent the 95% confidence interval around the mean. The dot colour represents cation concentration.**

4.2.2. Effect of Calcium Chloride Concentration on Total Cell Number in Spent and Fresh Media (8X10^6 cells/mL)

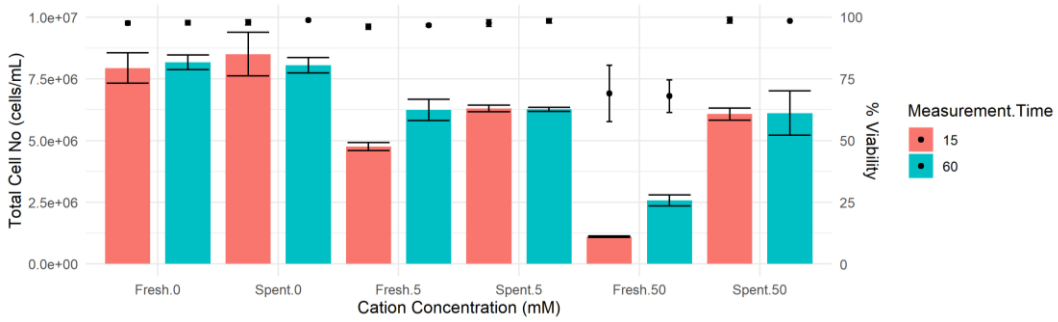


Figure 4.3: Total cell number (bars) and % viability (dots) after 15 and 60-minutes incubation of CHO K1 cells with calcium chloride. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in either the same media (spent) or new media (fresh) at 8X10^6 cells/mL. A 10X solution of calcium chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm prior to being sampled. Each bar/dot is the mean of 3 biological replicates and the error bars represent the standard deviation of those replicates.

Figure 4.4a **Figure 4.4b**

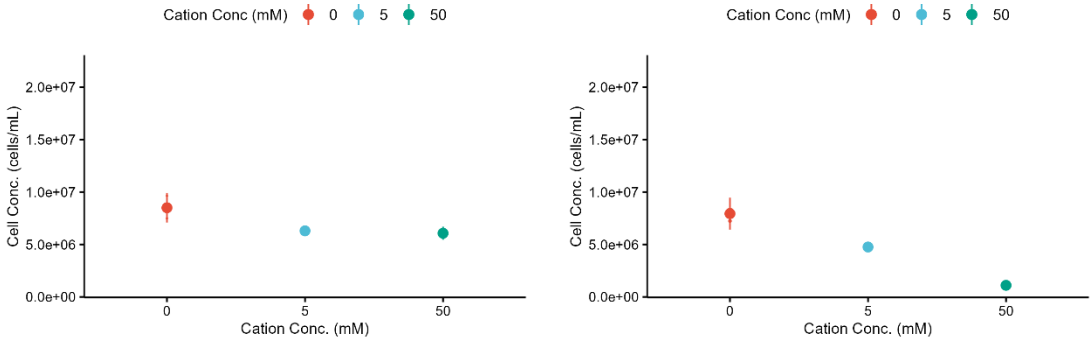


Figure 4.4: Total cell number remaining in solution measured after 15 minutes incubation with calcium chloride. Each dot is an individual biological replicate. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent, Figure 4.4a) or fresh media (fresh, Figure 4.4b) at 8X10^6 cells/mL. A 10X solution of calcium chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm. The bars on the plot represent the 95% confidence interval around the mean. The dot colour represents cation concentration.

The overall amount of variability in the data at 15 minutes and whether a significant effect was observed compared to the negative control was assessed the Kruskal Wallis test (p = 0.05, spent media and p = 0.04, fresh media). Even though the number of total cells/mL remaining in solution decreased with increasing cation concentration none of the tested concentrations resulted in a significant effect compared to the negative control in either fresh or spent media when pairwise comparisons were performed. Regarding cell viability, the use of $CaCl_2$ as a flocculating agent with $8X10^6$ cells/mL only impacted on cell viability at 50mM in fresh media.

4.2.3. Effect of Manganese Chloride Concentration on Total Cell Number in Spent and Fresh Media ($2X10^7$ cells/mL)

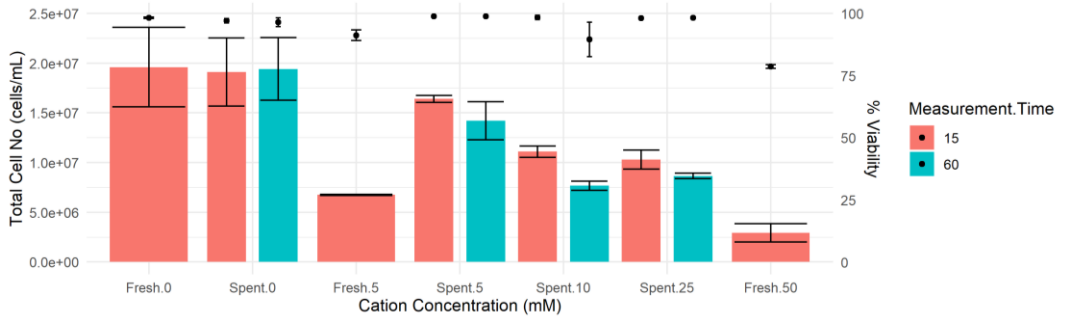


Figure 4.5: **Total cell number remaining in solution (bars) and % viability (dots) after 15 and 60-minutes incubation of CHO K1 cells with manganese chloride. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in either the same media (spent) or new media (fresh) at $2X10^7$ cells/mL. A 10X solution of manganese chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm prior to being sampled. The bars/dots representing the zero concentration are the mean of six biological replicates, all others are the mean of three biological replicates. The error bars represent the standard deviation of those replicates.**

Figure 4.6a **Figure 4.6b**

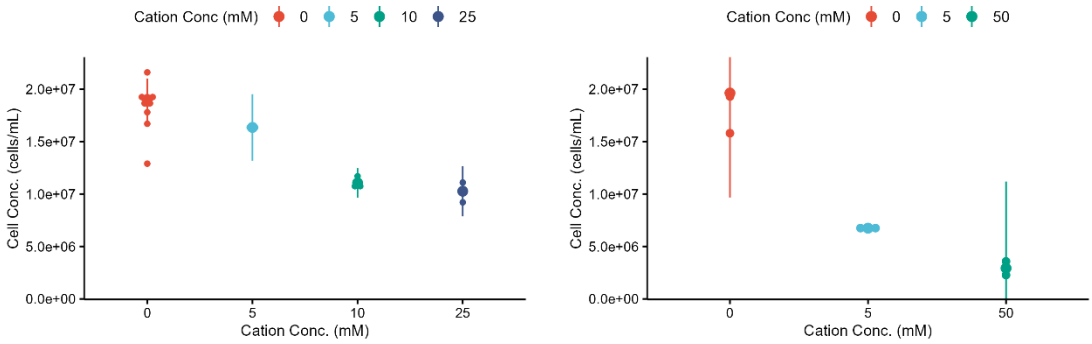


Figure 4.6: Total cell number remaining in solution measured after 15 minutes incubation with manganese chloride. Each dot is an individual biological replicate. Cells were grown overnight in Balan CD, centrifuged ad the cell pellet resuspended in the same media (spent, Figure 4.6a) or fresh media (fresh, Figure 4.6b) at 2X10^7 cells/mL. A 10X solution of manganese chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm prior to being sampled. The bars on the plot represent the 95% confidence interval around the mean. The dot colour represents cation concentration.

The overall amount of variation in the data and whether a significant effect was observed compared to the negative control was assessed the Kruskal Wallis test (p = 0.004, spent media and p = 0.04, fresh media). Even though the number of total cells/mL decreased with increasing cation concentration in both spent and fresh media none of the tested concentrations resulted in a significant effect compared to the negative control in either spent or fresh media.

In comparison to the data obtained when calcium chloride was used as the flocculating agent the use of manganese chloride resulted in a greater reduction in the total number of cells remaining in solution. For example, at a final concentration of 50mM, manganese chloride resulted in total cell clumping in spent media (the amount of clumping was so great the Cedex XS system couldn't analyse it) after 15 min compared to the statistically insignificant effect observed with 50mM calcium chloride. A similar effect was observed when comparing the effect of calcium and manganese on the number of cells remaining in suspension in fresh media at a final cation concentration of 50mM. In fresh media the total number of cells remaining in solution after 15 minutes treatment with 50mM calcium chloride was between 5X10^6 cells /mL and 1X10^7

cells/mL while the equivalent figure for 50mM manganese chloride was $<5X10^6$ cells/mL. It was also observed that 50mM manganese chloride had a greater impact on culture viability in fresh media than calcium chloride but as total cell number encompasses both viable and non-viable cells the impact of viability on the reduction in cell number was controlled for.

Given that the addition of manganese chloride to a final concentration of 50mM in spent media resulted in gross cell clumping it was decided to also investigate the effect of 25mM and 10mM manganese chloride on cell number in spent media only. While a reduction in the total number of cells remaining in solution after 15 minutes was observed this was not significant when compared to the negative control.

Regarding cell viability, the use of manganese chloride as a flocculating agent with $2X10^7$ cells/mL didn't impact on cell viability in spent media at any concentration after 15 minutes but did impact on cell viability at 5mM and 50mM after 15 minutes in fresh media. While the essential role manganese plays as an enzymatic co factor during metabolism is well established, there is evidence that at trace amounts it can be toxic. Francisco et al. (2021) used CHO cell lines to assess the cytotoxic and genotoxic effects of manganese concentration in groundwater in Brazil. They showed that as little as 1.5mg/L manganese had a statistically significant effect on CHO culture viability after twenty-four hours compared to the negative control. The concentration of manganese present in Balan CD (the media used in this study) is not known but given the high cell numbers used in this study it's possible that the level of manganese in the media is reduced sufficiently such that the addition of manganese chloride for flocculation was not toxic. In contrast, because the experimental design called for the pelleting and subsequent resuspension of cells in fresh media prior to the addition of manganese chloride for flocculation, cells resuspended in fresh media experienced a slightly higher concentration of manganese chloride which could have been sufficiently high enough to be toxic.

4.2.4. Effect of Manganese Chloride Concentration on Total Cell Number in Spent and Fresh Media (8X10^6 cells/mL)

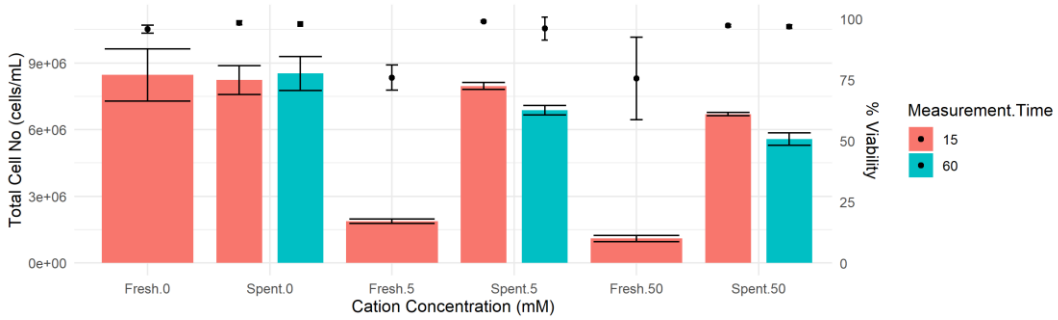


Figure 4.7: **Total cell number (bars) and % viability (dots) after 15 and 60-minutes incubation of CHO K1 cells with manganese chloride. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in either the same media (spent) or new media (fresh) at 8X10^6 cells/mL. A 10X solution of manganese chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm prior to being sampled. Each bar/dot is the mean of 3 biological replicates and the error bars represent the standard deviation of those replicates.**

Figure 4.8a **Figure 4.8b**

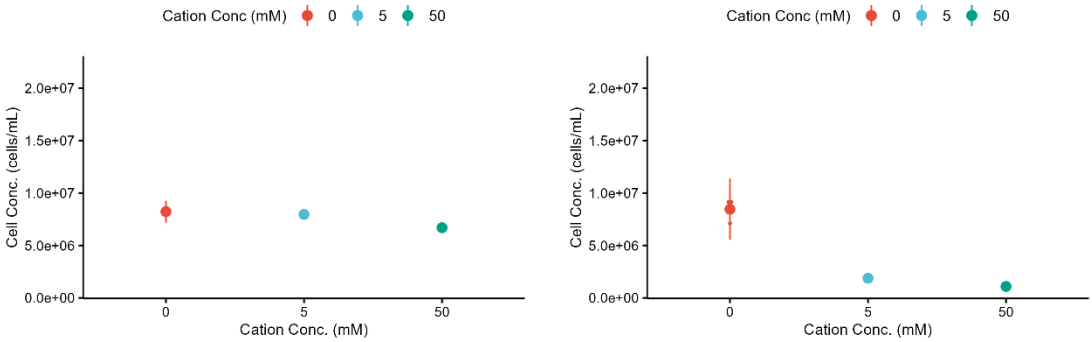


Figure 4.8: **Total cell number remaining in solution measured after 15 minutes incubation with manganese chloride. Each dot is an individual biological replicate. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent, Figure 4.8a) or fresh media (fresh, Figure 4.8b) at 8X10^6 cells/mL. A 10X solution of manganese chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm. The bars on the plot represent the 95% confidence interval around the mean. The dot colour represents cation concentration.**

In comparison to the effect observed when 50mM calcium chloride (Figure 4.3) was used as the flocculating agent a greater reduction in the number of cells remaining in solution after 15 minutes was observed with 50mM manganese chloride (Figure 4.7) and spent media compared to the negative control, while culture viability was similar regardless of which cation was used. Conversely, in fresh media a similar number of cells remained in solution after fifteen minutes incubation regardless of whether calcium or manganese was used (Figures 4.3 and 4.7). However, sixty minutes incubation with 50mM manganese chloride in fresh media resulted in gross cell clumping and no data was obtained (Figure 4.7) whereas under the same conditions with calcium chloride an increase in the number of cells remaining in solution was observed compared to the fifteen minutes time point (Figure 4.3). Incubation of CHO K1 cells with 5mM manganese in fresh media resulted in a larger reduction in the total number of cells remaining in suspension compared to the equivalent conditions for calcium. Indeed, the reduction in the total number of cells remaining in suspension following incubation with manganese in fresh media was similar for both 5mM and 50mM and no difference in the culture viability was observed (Figure 4.7).

The overall amount of variability in the data at 15 min and whether a significant effect was observed compared to the negative control was assessed the Kruskal Wallis test ($p = 0.05$, spent media and $p = 0.03$, fresh media). Even though the number of total cells/mL decreased with increasing cation concentration in both spent and fresh media none of the tested concentrations resulted in a significant effect compared to the negative control in either spent or fresh media.

The use of $MnCl_2$ as a flocculating agent with $8X10^6$ cells/mL had no impact on cell viability at any concentration in spent media after 15 minutes. However, the use of $MnCl_2$ as a flocculating agent impacted on cell viability at both 5mM and 50mM in fresh media after 15 minutes. This was the identical trend as seen with $2X10^7$ cells/mL.

4.2.5. Effect of Ferric Chloride Concentration on Total Cell Number in Spent and Fresh Media (2X10^7 cells/mL)

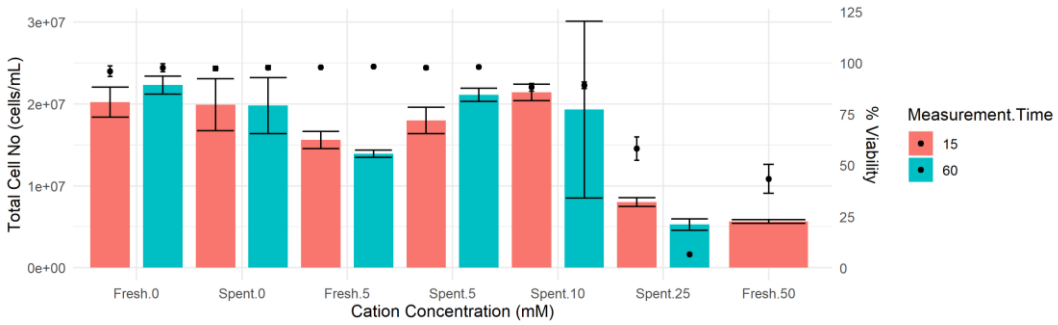


Figure 4.9: **Total cell number remaining in solution (bars) and % viability (dots) after 15 and 60-minutes incubation of CHO K1 cells with ferric chloride. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in either the same media (spent) or new media (fresh) at 2X10^7 cells/mL. A 10X solution of ferric chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm prior to being sampled. The bars/dots representing the zero concentration are the mean of six biological replicates, all others are the mean of three biological replicates. The error bars represent the standard deviation of those replicates.**

Figure 4.10a **Figure 4.10b**

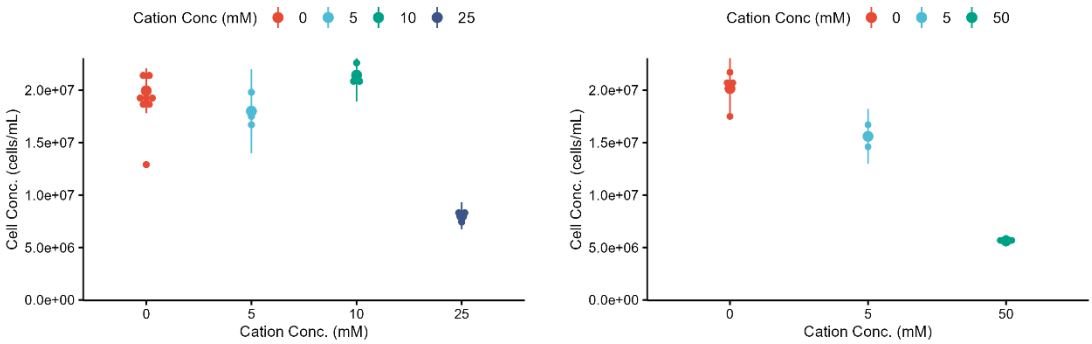


Figure 4.10: **Total cell number remaining in solution measured after 15 minutes incubation with ferric chloride. Each dot is an individual biological replicate. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent, Figure 4.10a) or fresh media (fresh, Figure 4.10b) at 2X10^7 cells/mL. A 10X solution of ferric chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm prior to being sampled. The bars on the plot represent the 95% confidence interval around the mean. The dot colour represents cation concentration.**

No data was plotted for the 50mM concentration in spent media at either time point in spent media or at 60 minutes in fresh media (Figure 4.9). This was because the amount of clumping in these samples was so great the Cedex XS system couldn't analyse them. Consequently, it was decided to investigate the effect of 25mM and 10mM ferric chloride on cell number in spent media only. The overall amount of variability in the data at 15 min and whether a significant effect was observed compared to the negative control was assessed the Kruskal Wallis test ($p = 0.07$, spent media and $p = 0.2$, fresh media). Even though the number of total cells/mL decreased with increasing cation concentration in fresh media (Figure 4.10b) and with 25mM ferric chloride in spent media (Figure 4.10a) none of the tested concentrations resulted in a significant effect compared to the negative control being observed when pairwise comparisons were performed.

Regarding cell viability, the use of ferric chloride as a flocculating agent with $2X10^7$ cells/mL had no impact on cell viability at 5mM in either spent or fresh media. 10mM, 25mM impacted on cell viability in spent media as did 50mM in fresh media.

4.2.6. Effect of Ferric Chloride Concentration on Total Cell Number in Spent and Fresh Media ($8X10^6$ cells/mL)

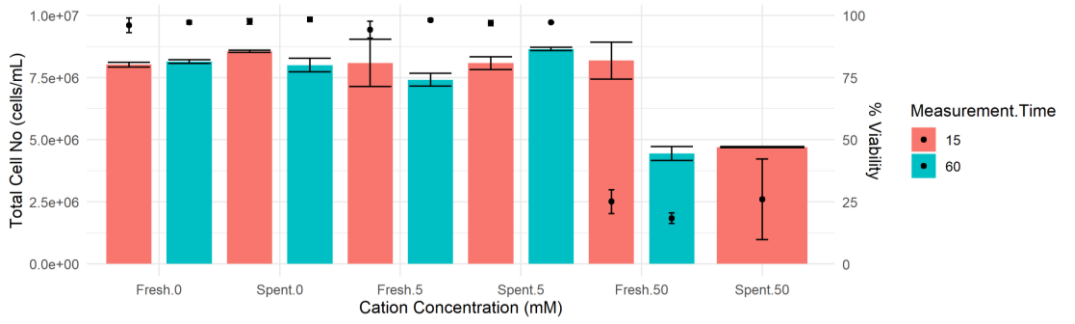


Figure 4.11: Total cell number (bars) and % viability (dots) after 15 and 60-minutes incubation of CHO K1 cells with ferric chloride. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in either the same media (spent) or new media (fresh) at $8X10^6$ cells/mL. A 10X solution of ferric chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm prior to being sampled. Each bar/dot is the mean of 3 biological replicates and the error bars represent the standard deviation of those replicates.

Figure 4.12a **Figure 4.12b**

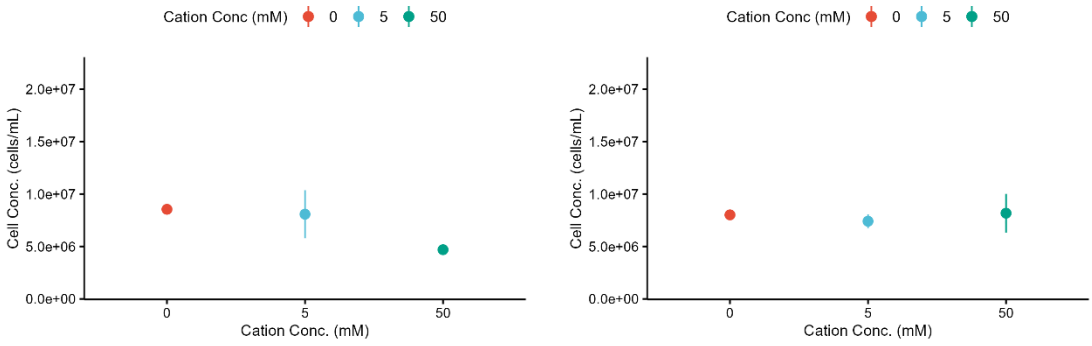


Figure 4.12: Total cell number remaining in solution measured after 15 minutes incubation with ferric chloride. Each dot is an individual biological replicate. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent, Figure 4.12a) or fresh media (fresh, Figure 4.12b) at 8X10^6 cells/mL. A 10X solution of ferric chloride was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm prior to being sampled. The bars on the plot represent the 95% confidence interval around the mean. The dot colour represents cation concentration.

No data was plotted for the 50mM concentration in spent media at 60 minutes. This was because the amount of clumping in these samples was so great the Cedex XS system couldn't analyse them. The overall amount of variability in the data at 15 min and whether a significant effect was observed compared to the negative control was assessed the Kruskal Wallis test (p = 0.07, spent media and p = 0.2, fresh media). Even though the number of total cells/mL decreased with increasing cation concentration in both spent and fresh media none of the tested concentrations resulted in a significant effect being observed compared to the negative control in spent media.

Regarding cell viability, the use of $FeCl_2$ as a flocculating agent with 8X10^6 cells/mL had no impact on cell viability at 5mM in both spent and fresh media at either time point. However, there was a significant impact on cell viability at 50mM in both spent and fresh media at both time points.

Overall, ferric chloride was most effective at reducing the number of CHO K1 cells remaining in solution followed by manganese chloride and calcium chloride under the conditions tested. Given the pH data presented in Table 4.2 that's not surprising as pH is known to impact the flocculation of CHO K1 cells. What was surprising, was that that

a greater reduction in the number of cells remaining in solution was observed in fresh media than spent media. Given that the idea behind using fresh and spent media was to attempt to model the effect of pH on flocculation this was not expected and would suggest that a combination of factors rather than just pH influence flocculation. As regards the lack of statistical significance this was possible due to the sample size. Power analysis could have been performed to help estimate the appropriate sample size but given that the desired outcome was gross cell flocculation as opposed to a 10% decrease in the number of cells remaining in solution after treatment this was deemed unnecessary. 50mM calcium chloride, 25mM manganese chloride and 25mM ferric chloride were chosen for further investigation.

4.2.7. Investigation of the Effect of Divalent Cation Concentration on The Sedimentation Rate of CHO K1 Cells

Based on the results presented in Figures 4.1 to 4.12b inclusive the impact of using 50mM calcium chloride, 25mM manganese chloride and 25mM ferric chloride on the sedimentation rate of CHO K1 cells in spent media was investigated. The experimental methodology was modified in these experiments. Cells were incubated with cation for 15 minutes as per the previous set of experiments. However, an additional ten minute sedimentation step at room temperature post incubation was added. The results are presented in Figures 4.13a – 4.13c inclusive. The overall amount of variability in the data and whether a significant effect was observed compared to the negative control was assessed by the Kruskal Wallis test at a $p \leq 0.05$ ($p = 0.0001$). Whether the difference in cell numbers between the individual concentrations and the negative control was significant was tested in pairwise comparisons by the Wilcoxon test at a $p \leq 0.05$.

Figure 4.13a **Figure 4.13b** **Figure 4.13c**

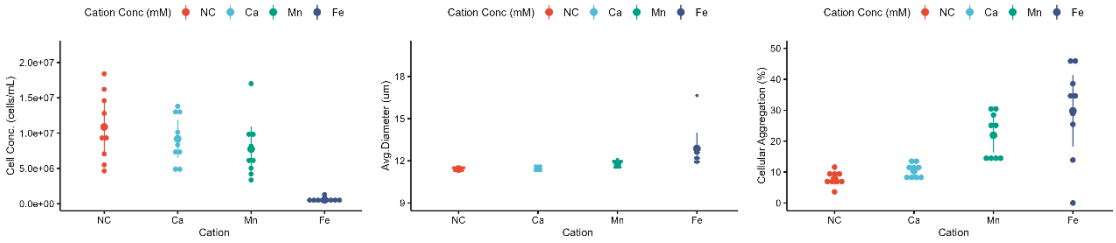


Figure 4.13: A) Total cell number remaining in solution after sedimentation.

B) Average diameter of cells remaining in solution after sedimentation.

C) % Aggregation of cells remaining in solution after sedimentation.

Each dot is an individual biological replicate. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent) at 2X10^7 cells/mL. A 10X solution of the appropriate cation was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37^{o}C, 170 rpm for 15 minutes followed by being left to settle out at room temperature for 10 minutes. The bars on the plot represent the 95% confidence interval around the mean. Cation concentration was as follows: Calcium chloride = 50mM, Manganese chloride = 25mM, Ferric chloride = 25mM.

Examination of Figure 4.13a revealed that the addition of ferric chloride at 25mM significantly increased the sedimentation rate of CHO K1 cells (p = 0.0003), as fewer cells remained in solution after sedimentation compared to the negative control (NC – cells settling under gravity). While the total number of cells/mL remaining in solution after treatment with 50mM calcium chloride and 25mM manganese chloride decreased, the impact on the sedimentation rate was not significant when compared to the negative control.

Use of the Cedex XS system enabled cell diameter to be measured. Examination of Figure 4.13b revealed that the addition of either manganese chloride or ferric chloride at 25mM significantly increased (p = 0.001, manganese chloride and p = 0.0002, ferric chloride) the cell diameter of CHO K1 cells compared to the negative control (NC – cells settling under gravity). There wasn't an increase in cell diameter for cells treated with 50mM calcium chloride compared to the negative control.

The Cedex XS also reports a parameter called "Aggregate Rate." This is a representation of what percentage of the cells imaged by the Cedex are not single cells. This parameter

was used in this study as a surrogate measurement of the degree to which the addition of cation to CHO K1 cells resulted in flocs being formed. Examination of Figure 4.13c revealed that the addition of manganese chloride and ferric chloride at 25mM have significantly increased ($p = 0.0003$, manganese and $p = 0.02$, ferric chloride) the % aggregation of CHO K1 cells compared to the negative control (NC – cells settling under gravity). While there appears to be an increase in the % aggregation of cells treated with 50mM calcium chloride this was not a significant effect compared to the negative control.

Taken as a whole, Figures 4.13a-4.13c inclusive indicate that the addition of either ferric chloride or manganese chloride at a final concentration of 25mM aids CHO K1 cells to settle out of suspension under gravity. It seems as if this is due to a combination of increased cell size and floc formation as measured by % aggregation.

The addition of calcium chloride at a final concentration of 50mM did not result in a statistically significant effect compared to the negative control for any of parameters measured.

4.2.8. Mimicking the Impact of Flocculation on a Producer Cell Line.

To aid the transfer of these observations from the lab to the "factory floor" the sedimentation experiments were repeated in the presence of spiked in antibody thereby mimicking the expression of a cloned recombinant protein. Rather than using a producer cell line which would have introduced questions regarding clonal variation and whether the results would hold true for all cell lines a human IgG antibody was purchased from a commercial supplier (Abcam) and spiked into cell culture broth ($2X10^7$ cells/mL at 1.5g/L) (Figure 4.14).

4.2.8.1. Experimental Approach

The experiment described in the previous section was repeated, i.e., addition of cation to cell culture broth, incubation for fifteen minutes, followed by cells being allowed to sediment for a further ten minutes, with three differences.

1. There is a body of literature on the ability of divalent metal cations such as iron (Ouellette et al., 2009), copper (Glover et al., 2015) and nickel (Krezel et al., 2010) to function as proteases and cleave proteins. A 10X EDTA stock was added to the cell culture broth immediately prior to sedimentation to chelate any free metal ions after flocculation had occurred and prevent/minimise this from happening. EDTA was added such that the final concentration of EDTA matched the cation concentration for ferric chloride (25mM) and manganese chloride (50mM). In order that the same volume of10X EDTA solution was added to all cations the final concentration of EDTA added to calcium chloride was 50mM. Addition of EDTA to the negative control allowed for the effect of dilution on total cell numbers/mL to be controlled for.
2. As the previous experiments with calcium and manganese were inconclusive the concentrations of calcium and manganese were modified. 100mM calcium and 50mM manganese were used to see if an increase in concentration resulted in a significant decrease in cell number compared to the negative control.
3. Sedimentation was allowed to occur at 37°C, 0 rpm rather than on the bench at room temperature.

The effect of cation addition on the sedimentation rate of cells was assessed by:

- the total number of cells/mL remaining in suspension after 10 minutes were counted. To disturb the sedimented cells as little as possible the suspended cells were sampled as 2 sequential 2mL fractions (EDTA upper and lower). The number of cells in each fraction were added together to provide the total number of cells remaining in suspension (Table 4.2).
- the sedimented cells were recovered by centrifugation and resuspended in a uniform volume (0.5mL) prior to counting the total number of cells/mL in the cell pellet.

To assess the impact of the variability in total cell number/mL in experiments performed on different days, the values obtained on three occasions for each cation and the negative control were averaged. The effect cation addition had on the cell broth pH prior to sedimentation was measured and is presented in Table 4.2. Both 25mM manganese chloride and 25mM ferric chloride affected pH.

	Run 1		Run 2		Run 3	
	pH	Temp (°C)	pH	Temp (°C)	pH	Temp (°C)
50mM Calcium Chloride	7.4	19.7	7.4	18.5	7.4	20.0
25mM Manganese Chloride	7.1	19.6	7.2	18.2	7.1	20.0
25mM Ferric Chloride	3.3	19.2	3.9	18.1	4.0	19.5
NC (Cells + PBS)	7.4	20.1	7.4	18.0	7.4	19.3

Table 4.2: **pH values for cell culture broth post spiking with different cation.**

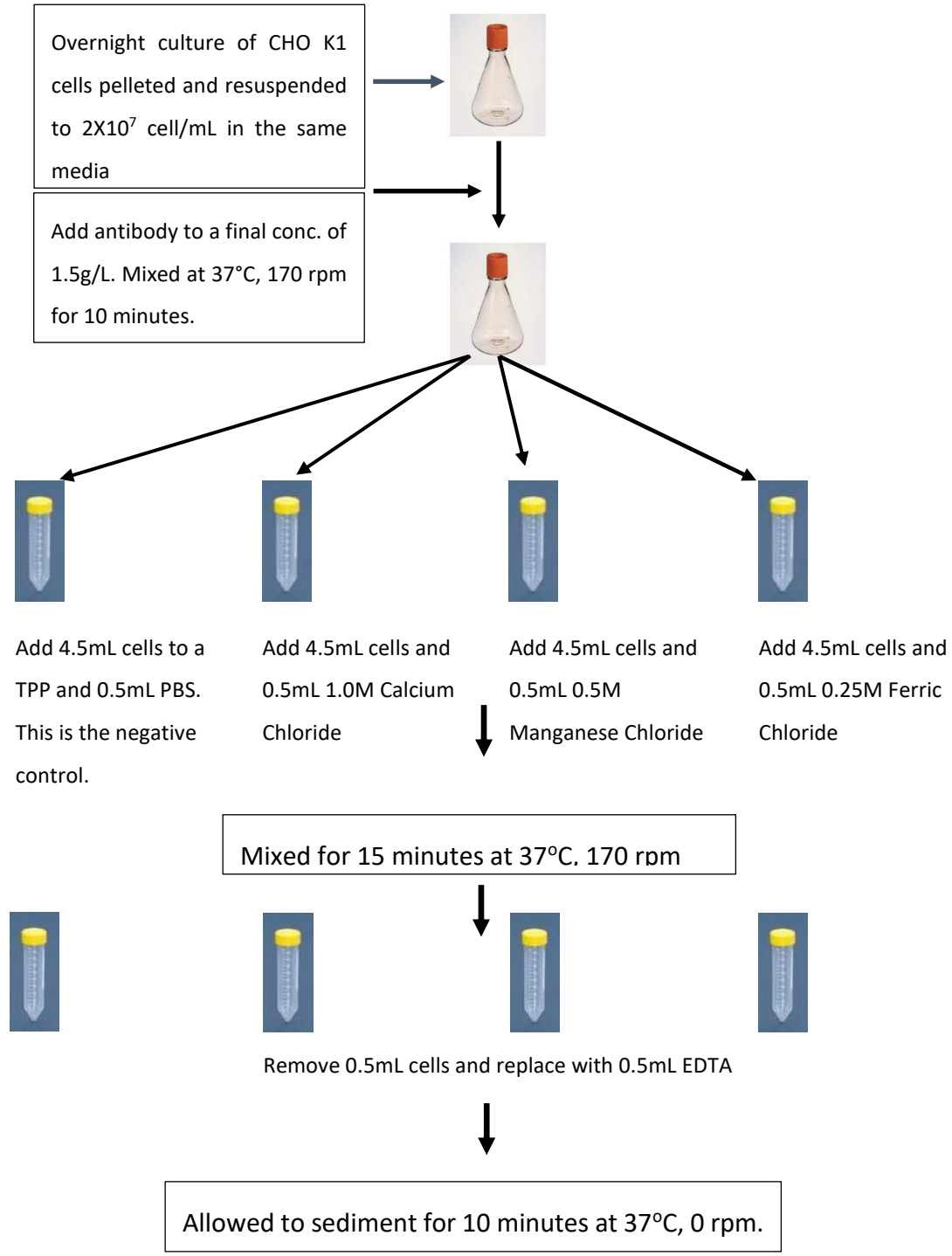


Figure 4.14: **Process flow diagram for the antibody spiking experiments.**

4.2.8.2. Effect of sedimentation on the number of total cells/mL remaining in suspension

The effect of cations on the total number of cells/mL is presented in Figures 4.15a and 4.15b and Table 4.3 and Table 4.4.

Examination of Figure 4.15a and Table 4.3 revealed that 100mM calcium chloride and 50mM manganese chloride caused a small reduction in the total number of cells/mL compared to the negative control after 15 min incubation with cation. Visually 25mM ferric chloride resulted in a significant reduction in the total number of cells/mL remaining in solution compared to the negative control. Figure 4.16 shows a distinct ring of cells deposited on the inner surface of the TPP tube containing 25mM ferric chloride. This ring of cells was always present in the TPP tube containing cells treated with 25mM ferric chloride and always absent in all other conditions investigated. However, once the effect of multiple pairwise comparisons on the p value was accounted for (Benjamini Hochberg correction) the effect was not statistically significant (p= 0.07).

Table 4.3 also revealed that the negative control showed a reduction in the number of cells/mL after 15 minutes incubation. This finding is neither surprising nor a cause for concern. Cation was added as a 10X stock and therefore to account for this dilution effect in the negative control, it was diluted by addition of the same volume of PBS. Diluting the negative control in this way allowed an assessment of how much of the reduction in cell number was due to the addition of cation as opposed to the dilution effect. However, it should be noted that the size of the reduction in cell number (26.3%) is greater than that expected due to dilution alone. The reason for this is not clear but it's for this reason that the reduction in cell number in Table 4.3 is compared to the measured number of cells/mL in the negative control after incubation rather than the starting number.

Comparison	Cells/mL (Target)	Cells/mL (Actual)	% Reduction	p Value	p adjusted
NC v Calcium Chloride	$1.47X10^7$	$1.37X10^7$	-6.9%	0.3	0.7
NC v Manganese Chloride	$1.47X10^7$	$1.30X10^7$	-12.0%	0.3	0.5
NC v Ferric Chloride	$1.47X10^7$	$1.01X10^7$	-31.7	0.01	0.07

Table 4.3: **% Reduction in Total Cell Number/mL remaining in solution post incubation at 37 °C, 170 rpm with different cations for 15 min. Adjusted p value is the p value following application of the Benjamini-Hochberg formula to the calculated p value**

Comparison	Cells/mL ($X10^7$)	Cells/mL ($X10^7$) Remaining in Suspension			% Reduction	P value	P adjusted
	Target	EDTA Upper Fraction	EDTA Lower Fraction	Combined Fractions			
Starting Flask v NC	1.47	0.496	0.623	1.12	-23.8	NA	NA
NC v Calcium chloride	1.12	0.264	0.325	0.588	-47.4	$4.1X10^{-5}$	$2.5X10^{-4}$
NC v Manganese chloride	1.12	0.165	0.338	0.503	-55.1	$4.1X10^{-4}$	0.002
NC v ferric chloride	1.12	0.394	0.552	0.946	-15.5	0.12	0.27

Table 4.4: **% Reduction in total number of cells/mL remaining in solution after the addition of EDTA and cells were allowed to settle out for 10 min at 37°C, 0 rpm. Adjusted p value is the p value following application of the Benjamini-Hochberg formula to the calculated p value.**

Figure 4.15a **Figure 4.15b**

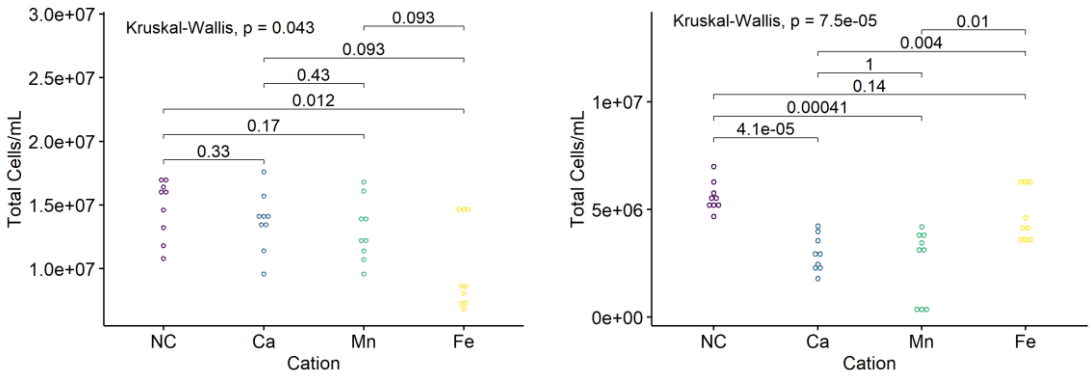


Figure 4.15: **a) Total cell number measured after 15 minutes incubation with various cations.**

b) Total cell number remaining in solution after cells were allowed to settle out.

Each dot is a technical replicate. The experiment was repeated on three separate occasions resulting in nine technical and three biological replicates for each condition. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent) at 2X10^7 cells/mL. A 10X solution of the appropriate cation was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm followed by cells being left to settle out at 37°C, 0 rpm for 10 minutes. The numbers on the plot are the p value after performing a Wilcoxon test prior to performing a Benjamini Hochberg adjustment for multiple comparisons. *Cation concentration was as follows: Calcium = 100mM, Manganese = 50mM, Ferric = 25mM.*

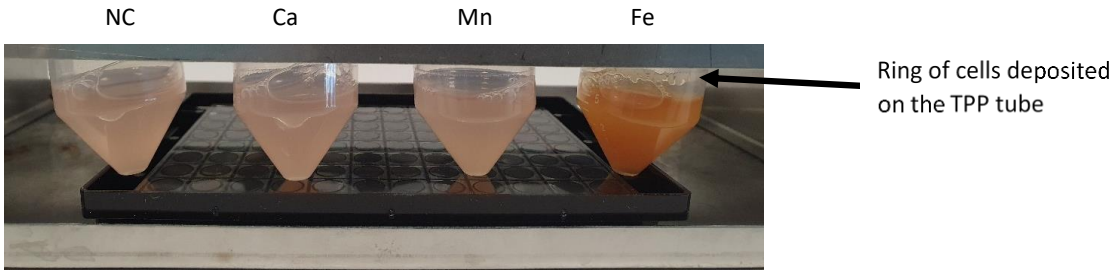


Figure 4.16: **Photo of TPP tubes post incubation at 37°C with different cations for 15 min showing the presence of a ring of clumped cells in the tube containing 25mM ferric chloride.**

The finding that only ferric chloride had a significant effect in reducing the total number of cells in suspension after 15 minutes incubation with cation did not apply once EDTA was added to the cells and the cells allowed to sediment for a further 10 minutes (post

sedimentation stage). Examination of Figure 4.15b and Table 4.4 revealed that 100mM calcium chloride (p = 0.0003) and 50mM manganese chloride (p=0.0002) resulted in a significant reduction in the total number of cells/mL remaining in solution, compared to the negative control, after being allowed to settle out for 10 minutes. While 25mM ferric chloride did result in a further reduction in cell numbers post sedimentation this was not significant when compared to the reduced cell numbers observed in the negative control.

4.2.8.3. Cell Viability and Measurement of Lactate Dehydrogenase Activity in the Cell Culture Media

It was observed that cell viability decreased along with the number of total cells/mL post incubation with 100mM calcium chloride, 50mM manganese chloride and 25mM ferric chloride. To assess whether cell lysis was responsible for the decrease in cell numbers lactate dehydrogenase activity in the media was measured. A compromised cell membrane is known to be correlated with decreased cellular viability. This allows for lactate dehydrogenase to leak out of the cell and into the media. Hence lactate dehydrogenase activity in the media can be used as an orthogonal method for measuring cell viability. A large increase in lactate dehydrogenase activity could be indicative of wholescale cell lysis.

The average % viability of cells treated with 100mM calcium chloride for 15 minutes was 94.3% (n=3). Average LDH activity decreased from 207 mUnits/mL in media from the negative control (98.4% viable) to 145 mUnits/mL in media from cells treated with 100mM calcium chloride. Allowing the cells to sediment for 10 minutes further decreased the cell viability to an average of 38.5% (n=3). This was accompanied by a decrease in average LDH activity from 318 mUnits/mL in media from the negative control (85.7%) to 132 mUnits/mL.

The average % viability of the cells treated with 50mM manganese chloride was 85.8% (n=3). This was accompanied by a decrease in average LDH activity from 207 mUnits/mL in media from the negative control to 53.3 mUnits/mL in media from cells treated with 50mM manganese chloride. Allowing the cells to sediment for 10 minutes further decreased the cell viability to an average of 75.1% (n=3). This was accompanied by a decrease in average LDH activity from 318 mUnits/mL in media from the negative

control to 16.7 mUnits/mL in media from cells treated with 50mM manganese chloride. However, except for one replicate from each of the post cation and post sedimentation samples manganese did not impact on % viability compared to the negative control.

The average % viability of the cells treated with 25mM ferric chloride was 49.7% (n=3). This was accompanied by a decrease in average LDH activity from 207 mUnits/mL in media from the negative control to 5 mUnits/mL in media from cells treated with 25mM ferric chloride. Allowing the cells to sediment for a further 10 minutes resulted in an increase in cell viability to an average of 67.7% (n=3). This was accompanied by a decrease in average LDH activity from 318 mUnits/mL in media from the negative control compared to 4.9 mUnits/mL in media from cells treated with 25mM ferric chloride.

All treatments resulted in an increase in cell diameter compared to the negative control post incubation with cation (Figure 4.17a). However, once the effect of multiple pairwise comparisons on the p value was accounted for (Benjamini Hochberg correction) only ferric chloride resulted in a statistically significant effect ($p = 0.0003$). Only ferric chloride ($p = 0.001$) and manganese chloride ($p = 0.0003$) had a significant impact on the average diameter of cells remaining in solution after 10 minutes sedimentation (Figure 4.17b) compared to the negative control. Comparing Figures 4.17a and 4.17b it appears as if there is a trend towards the cell diameter being smaller post sedimentation.

4.2.8.4. Effect on Cell Size

The effect of cations on average cell diameter is presented in Figures 4.17a and 4.17b.

Figure 4.17a

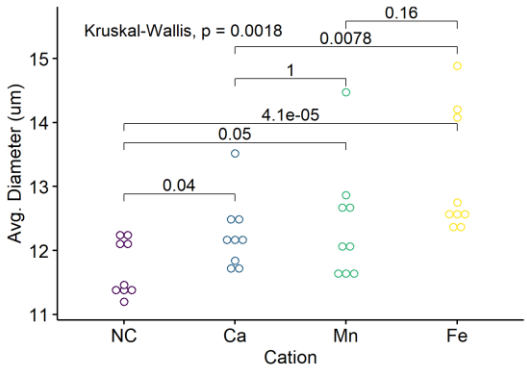


Figure 4.17b

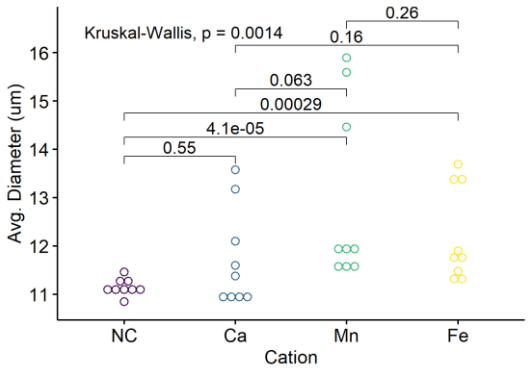


Figure 4.17 **a) Average cell diameter measured after 15 minutes incubation with multiple cations.**

b) Average cell diameter measured after cells were allowed to settle out.

Each dot is a technical replicate. The experiment was repeated on three separate occasions resulting in nine technical and three biological replicates for each condition. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent) at 2X10^7 cells/mL. A 10X solution of the appropriate cation was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37^{o}C, 170 rpm followed by cells being left to settle out at 37^{o}C, 0 rpm for 10 minutes. The numbers on the plot are the p value after performing a Wilcoxon test prior to performing a Benjamini Hochberg adjustment for multiple comparisons. *Cation concentration was as follows: Calcium chloride= 100mM, Manganese chloride= 50mM, ferric chloride = 25mM.*

4.2.8.5. Effect on Cellular Aggregation

An orthogonal measurement for assessing the effect of cation addition on cellular flocculation was % cellular aggregation. The effect of cation addition to cell broth on cellular aggregation is presented in Figures 418a and 418b.

Figure 4.18a

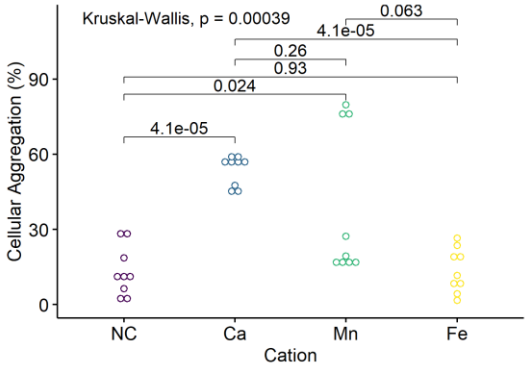


Figure 4.18b

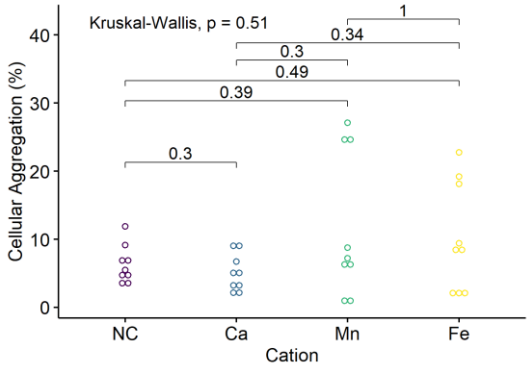


Figure 4.18: **a) % cellular aggregation measured after 15 minutes incubation with multiple cations.**

b) % cellular aggregation measured after cells were allowed to settle out.

Each dot is a technical replicate. The experiment was repeated on three separate occasions resulting in nine technical and three biological replicates for each condition. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent) at 2X10^7 cells/mL A 10X solution of the appropriate cation was added to the cells to achieve the indicated cation concentration and the cells were incubated at 37°C, 170 rpm followed by cells being left to settle out at 37°C, 0 rpm for 10 minutes. The numbers on the plot are the p value after performing a Wilcoxon test prior to performing a Benjamini Hochberg adjustment for multiple comparisons. *Cation concentration was as follows: calcium chloride= 100mM, manganese chloride = 50mM, ferric chloride = 25mM.*

Incubating cells with 100mM calcium chloride or 50mM manganese chloride resulted in an increase in cellular aggregation compared to the negative control post incubation with cation (Figure 4.18a). 25mM ferric chloride did not increase cellular aggregation. Given the previous observation (Figure 4.16) about the presence of a ring of cells deposited on the surface of the TPP tube containing these cells this result is not unexpected. However, once the effect of multiple pairwise comparisons on the p value was accounted for (Benjamini Hochberg correction) only calcium chloride resulted in a statistically significant effect (p = 0.0003) compared to the negative control. Post sedimentation there was no significant effect on cellular aggregation, for cells remaining

in solution, between cells exposed to cation and the negative control. Indeed, the percentage of cells which were aggregated declined for all treatments.

4.2.8.6. Effect of Cation Addition on Antibody Concentration (g/L) and Structure

As well as measuring the effect of cation addition on total cell number and cell size it was necessary to assess the effect adding divalent cations had on the yield and structure of the recombinant protein. A human IgG was purchased and spiked into CHO K1 cells at 1.5g/L. TPP tubes containing the cells, antibody and cation were sampled before cation addition, after 15 minutes incubation in the presence of cation and after sedimentation occurred (Section 2.8.1). An ELISA assay specific for human IgG was used to measure antibody concentration or yield at each point (Section 2.9). Post sedimentation the supernatant was further clarified by centrifugation and the antibody purified using Protein A/G spin columns (Section 2.8.3). SDS-PAGE (both reducing and non-reducing) (Section 2.6.3.1 and Section 2.6.3.2 and size exclusion chromatography (Section 2.8.4) were used to investigate if any changes had occurred in the structure of the recombinant protein.

The overall amount of variability in the data and whether it differed significantly from the negative control at each sampling point was assessed by the Kruskal Wallis test at a $p \leq 0.05$. Pairwise comparisons between individual treatments were performed using a Wilcoxon test at a $p \leq 0.05$. The effect of cation addition on antibody concentration at each sampling point are presented in Figures 4.19a and 4.19b. The effect of cation addition on antibody yield post Protein A/G purification is presented in Figure 4.19c.

A reduction in antibody concentration from cells treated with 25mM ferric chloride at both the post cation and sedimentation stages compared to the negative control was observed. This was not found to be statistically significant at $p \leq 0.05$. The 100mM calcium chloride or 50mM manganese chloride treatments did not lead to a reduction in antibody compared to the negative control.

While there was a reduction in the yield of purified antibody (Figure 4.19c) from cells treated with cation compared to the negative control it was not significant for any of the tested cations. The average yield of antibody following treatment with the various

cations compared to the negative control was 45% for manganese chloride, 53% for ferric chloride and 53% for calcium chloride respectively.

Figure 4.19a **Figure 4.19b** **Figure 4.19c**

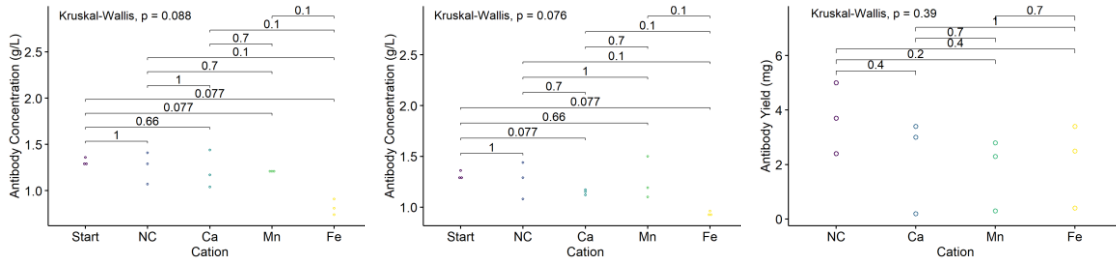


Figure 4.19: **a) Antibody concentration post cation incubation (g/L).**

b) Antibody concentration post sedimentation (g/L).

c) Antibody yield (grams antibody) post purification.

Each dot is an individual biological replicate. Antibody concentration was measured in cell free supernatant by an ELISA specific for human IgG using the same antibody as a standard in the ELISA. The numbers on the plot are the p value after performing a Wilcoxon test prior to performing a Benjamini Hochberg adjustment for multiple comparisons. *Cation concentration was as follows: calcium chloride= 50mM, manganese chloride= 25mM, ferric chloride = 25mM.*

Examination of Figure 4.19c shows that there was an issue with the purification for one of the replicates of each cation treatment. The purification protocol utilised called for the clarified cell culture supernatant to be diluted with an equal volume of buffer and this step was performed for replicate three a white precipitate was observed. When these values are removed from the yield calculations the average yield of antibody compared to the negative control for each of the cation treatments is 62% for manganese chloride, 72% for ferric chloride and 76% calcium chloride respectively.

4.2.8.7. Analysis of antibody fragmentation patterns following incubation with divalent cations

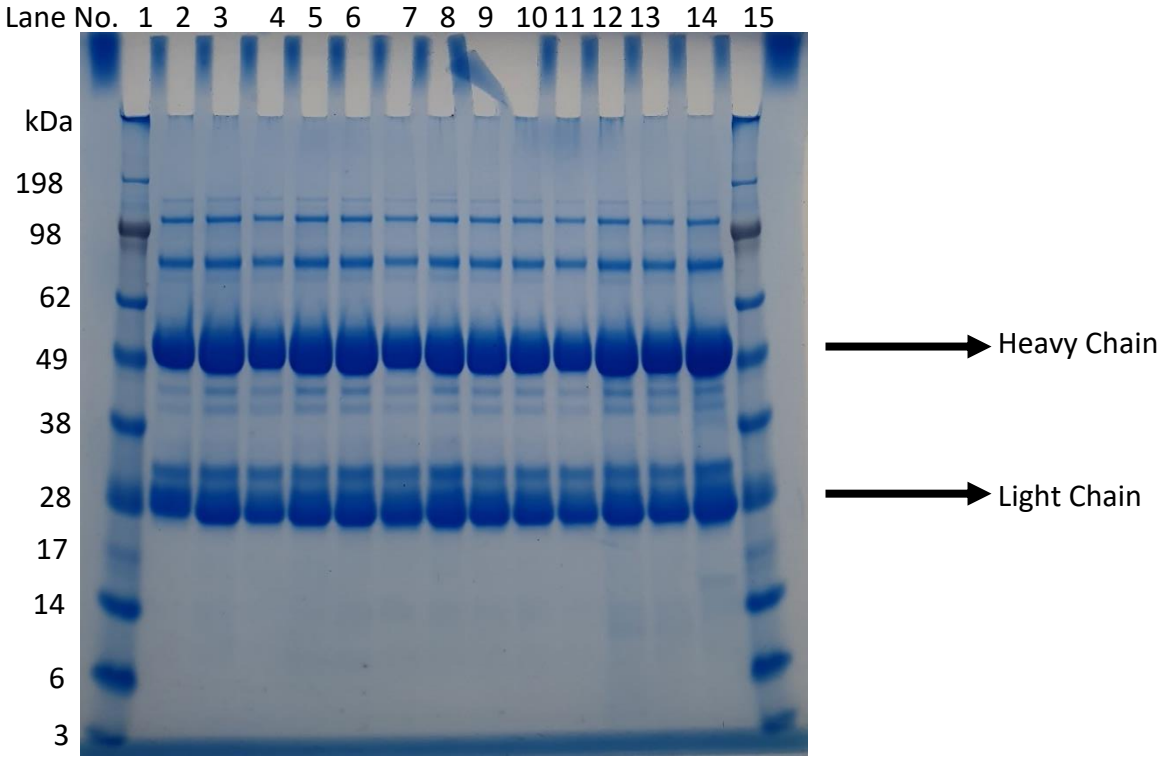
It has been reported in the literature (Yashiro et al., 2003*)* that metal cations such as iron can function as proteases and cleave proteins containing disulphide bonds. Antibodies fall into this class of proteins.

Therefore, to investigate and quantify if this phenomenon had occurred the protein A/G eluates were analysed by reducing and non-reducing SDS PAGE (Figures 4.20 and 4.21 respectively) and by size exclusion chromatography (SEC-HPLC). This allowed for an

estimate to be made of the molecular weight of the purified antibody and the degree of fragmentation to be quantified. An aliquot of the antibody used to spike the CHO K1 culture was included on the gels as a comparison.

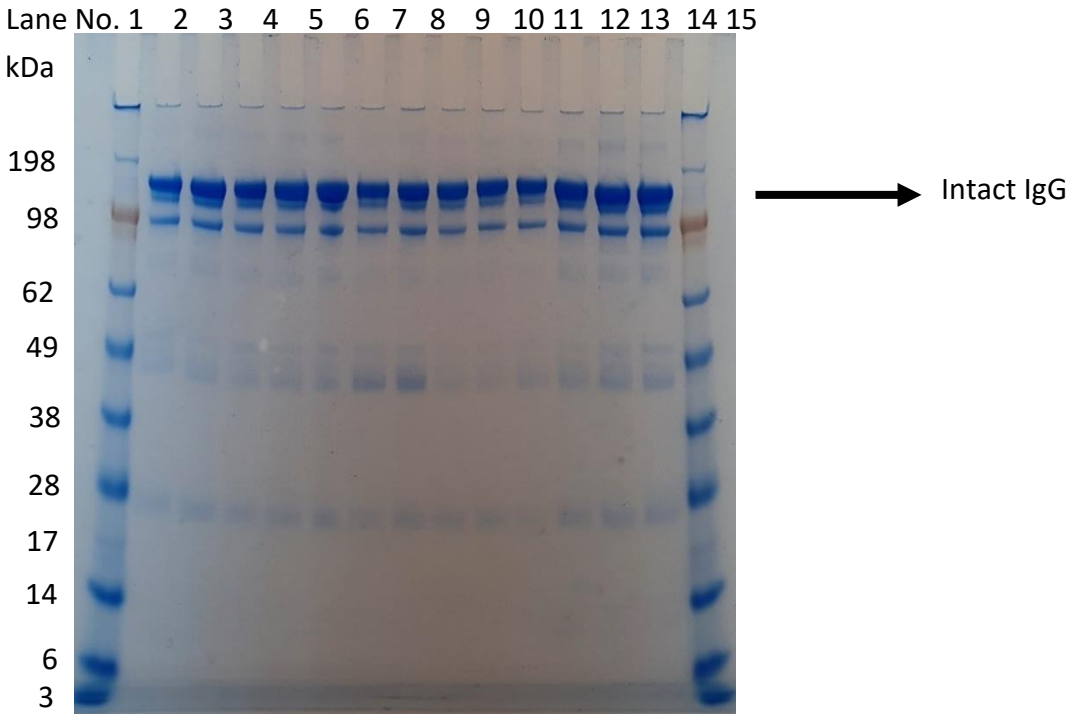
The eluates from each cation were analysed on a 4-12% gradient reducing and non-reducing SDS gels (Figures 4.20 and 4.21) and visually compared. There are more bands present overall than would be expected following reduction with DTT. This was queried with the supplier who advised that "*Those bands are likely the result of aggregation of the light and heavy chains. We do not believe those bands are impurities as this product showed single arc on IEP against antisera to whole human serum."* An aliquot of the antibody prior to spiking was included as a comparator. Visually no difference was observed in the number of bands detected in the calcium and manganese eluates compared to the comparator (Figure 4.20, lane 8) or the negative control on the reducing gel. There are some differences in staining intensity of the bands, but these have not resulted in extra bands being detected in any lane. There appears to be some extra low molecular weight bands (< 17kDa) in the ferric chloride eluates compared to the comparator and negative control.

Visually no difference was observed in the number of bands detected in the calcium eluates compared to the comparator (Figure 4.21, lane 8) or the negative control on the non-reducing gel. There are some differences in staining intensity of the bands, but these have not resulted in extra bands being detected in the lanes corresponding to these eluates. The banding pattern of the manganese eluates show a difference to the negative control and the comparator. There is a band present at approximately 98kDa in the comparator and negative control which is missing in the manganese. The banding pattern of the ferric chloride eluates show a difference to the negative control and the comparator. There is a band present above the 198kDa which is absent in the comparator and untreated eluates.



Lane No	Sample
1	**Molecular weight ladder**
2	**Negative control (Pr A Purified N=1)**
3	**Negative control (Pr A Purified N=2)**
4	**Negative control (Pr A Purified N=3)**
5	**Treated with 100mM $CaCl_2$ (Pr A Purified N=1)**
6	**Treated with 100mM $CaCl_2$ (Pr A Purified N=2)**
7	**Treated with 100mM $CaCl_2$ (Pr A Purified N=3)**
8	**Comparator – Commercial IgG (Abcam)**
9	**Treated with 50mM $MnCl_2$ (Pr A Purified N=1)**
10	**Treated with 50mM $MnCl_2$ (Pr A Purified N=2)**
11	**Treated with 50mM $MnCl_2$ (Pr A Purified N=3)**
12	**Treated with 25mM $FeCl_2$ (Pr A Purified N=1)**
13	**Treated with 25mM $FeCl_2$ (Pr A Purified N=2)**
14	**Treated with 25mM $FeCl_2$ (Pr A Purified N=3)**
15	**Molecular weight ladder**

Figure 4.20: **Reduced SDS_PAGE analysis of antibody following Pr A/G Purification.** *Each lane is a biological replicate. Cation concentration was as follows: calcium chloride= 100mM, manganese chloride = 50mM, ferric chloride = 25mM.*



Lane No	Sample
1	**Molecular weight ladder**
2	**Negative control (Pr A Purified N=1)**
3	**Negative control (Pr A Purified N=2)**
4	**Negative control (Pr A Purified N=3)**
5	**Treated with 100mM $CaCl_2$ (Pr A Purified N=1)**
6	**Treated with 100mM $CaCl_2$ (Pr A Purified N=2)**
7	**Treated with 100mM $CaCl_2$ (Pr A Purified N=3)**
8	**Comparator – Commercial IgG (Abcam)**
9	**Treated with 50mM $MnCl_2$ (Pr A Purified N=1)**
10	**Treated with 50mM $MnCl_2$ (Pr A Purified N=2)**
11	**Treated with 50mM $MnCl_2$ (Pr A Purified N=3)**
12	**Treated with 25mM $FeCl_2$ (Pr A Purified N=1)**
13	**Treated with 25mM $FeCl_2$ (Pr A Purified N=2)**
14	**Treated with 25mM $FeCl_2$ (Pr A Purified N=3)**
15	**Molecular weight ladder**

Figure 4.21: **Non-Reduced SDS_PAGE analysis of antibody following Pr A/G Purification.** *Each lane is a biological replicate. Cation concentration was as follows: calcium chloride= 100mM, manganese chloride = 50mM, ferric chloride = 25mM.*

4.2.8.8. Size Exclusion HPLC Analysis of Purified Antibody

Figure 4.22a

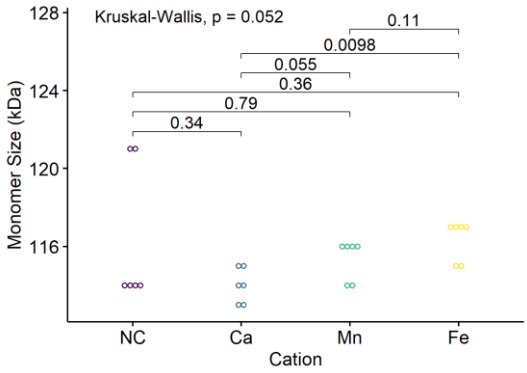


Figure 4.22b

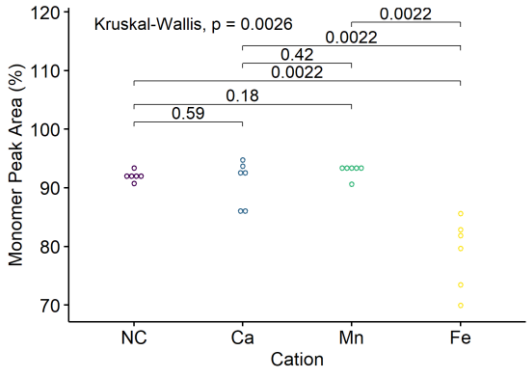


Figure 4.22c

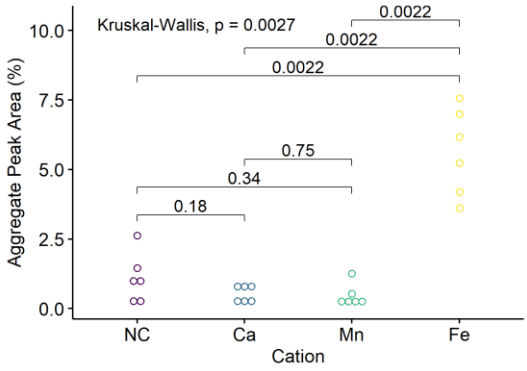


Figure 4.22d

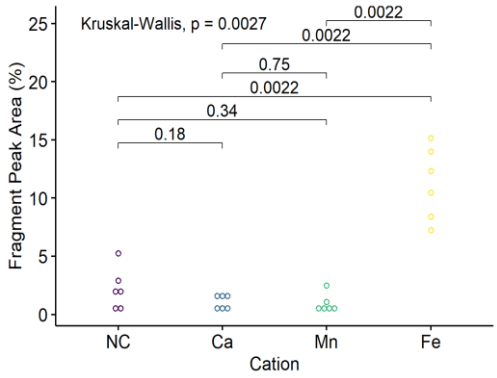


Figure 4.22: **a) Molecular weight (kDa) of purified antibody monomer.**

b) % area of purified antibody monomer.

c) % area of purified antibody aggregates.

d)% area of purified antibody fragments.

Each dot is a technical replicate. Each sample was analysed twice resulting in six technical and three biological replicates for each condition. Samples (50 mL) were assayed on an Agilent 1100 HPLC system with a VWD detector at 280nm and a quaternary pump using a TOSOH TSKgel G3000SWxl 30 cm column with an internal diameter of 7.8mm and a particle size of 5mm. The mobile phase was 100mM phosphate buffer pH8.0 running at 1.0mL/min. *Cation concentration was as follows: Cation concentration was as follows: calcium chloride= 100mM, manganese chloride = 50mM, ferric chloride = 25mM.*

There were no significant changes in antibody size (Figure 4.22a) post protein A/G purification for 100mM calcium chloride, 50mM manganese chloride or 25mM ferric

chloride compared the negative control (cells treated with PBS). There is a discrepancy between the calculated size (~116 kDa) and the theoretical size of an IgG molecule (150 kDa). This has been attributed to measurement uncertainty in the regression line used to calculate molecular weight.

There was a significant effect on the monomer % area (Figure 4.22b) for antibody purified from cells treated with 25mM ferric chloride ($p = 0.013$) compared to the negative control. There was no significant effect compared to the negative control on antibody purified from cells treated with either calcium chloride or manganese chloride.

The total aggregate % area (Figure 4.22c) was calculated by adding the % area of all the individual peaks with a molecular weight larger than that of the monomer. The was a significant effect with regards to % aggregation for antibody purified from cells treated with 25mM ferric chloride ($p = 0.013$) compared to negative control. There was no significant effect related to antibody aggregation between antibody purified from cells treated with either calcium chloride or manganese chloride and the negative control.

The total fragment % area (Figure 4.22d) was calculated by adding the % area of all the individual peaks with a molecular weight smaller than that of the monomer. There was no significant effect on the peak area associated with antibody fragments for any of the cation treatments compared to the negative control. Taken as a whole the GPC results mirror those of the non-reduced SDS PAGE analysis.

4.3. Discussion

Analysis of cell size, cell aggregation and LDH activity data is confined to experiments on $2X10^7$cells/mL in spent media as these most closely replicate industrial processes. Analysis of the effect of media and cation concentration on the ability of CHO cells to flocculate encompasses cells at both $2X10^7$cells/mL and $8X10^6$cells/mL.

Flocculation aims to reduce the size distribution of intact cells and cell debris present in cell culture broth. However, two particles with a similar overall surface charge, such as cells, will repulse each other. Therefore, for charged particles to flocculate this property needs to be overcome. Flocculants achieve this in one of two ways. They either possess a charge and thus attract particles of the opposite charge causing them to precipitate out of solution or they alter the ionic strength of the fluid they are suspended in. The expectation was that the addition of divalent cations would increase in the ionic

strength of the cell broth, compressing the electric double layer and allowing attractive van der Waal forces to predominate and hence cells would flocculate.

There are two potential explanations for the reduction in cell numbers observed post cation addition.

- The addition of cation led to cell death and cell lysis.
- The addition of cation promoted cellular flocculation and therefore the number of single cells present in the sample declined.

To discover which scenario best explains the data lactate dehydrogenase (LDH) activity in the media was measured as a surrogate for cell lysis. This approach was based on the fact that as cell die the cell membrane becomes compromised and the contents of the cell leak out into the media. Therefore, large scale cell lysis should lead to increased LDH activity in the media.

The data from the LDH assays appears confused. As expected, a reduction in cell viability was associated with an increase in LDH activity in the media for both the negative control. However, this was not the case for cells treated with any of the cations studied. In all cases the LDH activity was lower in the cation treated cells than the negative control. This finding could be explained by the results of Bornhorst et al. (2012) and Kaur et al (2017) who demonstrated that manganese reduces the levels of NADH intracellularly in a human lung cell line and *E. coli* respectively. NADH is a cofactor of the enzyme LDH and a reduction in the amount of NADH available in the media could impact on the activity of LDH. Similarly, the lower levels of LDH activity in media from cells treated with 25 mM ferric chloride can be explained by the finding of Lvovich and Scheeline (1995) that iron ions can form complexes with NADH. By forming complexes with NADH, iron reduces the amount of NADH available and hence the activity of LDH. This is supported by the finding of Oexle et al. (1999) that LDH activity is inversely correlated with iron concentration.

Since the LDH assay data was inconclusive with regards to the question of whether cells were sedimenting or lysing a cell balance was performed for each cation treatment. This was performed by summing the total number of cells at each step (post cation, post sedimentation and in the cell pellet) and reporting that as a percentage of the expected total number of cells. The expected total number of cells in each TPP tube was $1X10^8$

cells (5mL @ $2X10^7$ cells/mL). The average cell balance across the three replicate experiments ranged from 96% for ferric chloride to 126% for the negative control. These numbers support the conclusion that the reduction in cell numbers was due to sedimentation and not cell lysis.

4.3.1. Effect of pH

Westoby et al. (2011) showed that 10mM calcium chloride was sufficient to achieve flocculation at low pH. As it's generally accepted that cells grown in batch mode in flasks with no pH control, pH decreases over time in cell culture as metabolic by products accumulate in the media the experiments presented here modelled the effect of pH by centrifuging an overnight culture of CHO cells and resuspending the cells in fresh media prior to the addition of cation.

The results presented here for all cations demonstrate that the reduction in the total number of cells/mL after 15 minutes incubation with cation was greater in fresh media (high pH) than spent (low pH) media. The pH of Balan CD media is typically in the range 6.8-7.2. At this pH cells have an overall negative surface charge due to the presence of sialic acid groups on the cell surface and would be attracted to positively charged objects such as calcium, manganese, and iron ions via charge bridging. Additionally, it's well recognised in colloidal science that the addition of salt will modify a solutions ionic strength and reduce the thickness of the electric double layer. Consequently the repulsive forces which keep similarly charged cells apart in solution are no longer sufficiently strong enough to overcome attractive van der Waal forces, resulting in flocculation.

Iron (III), in the form of ferric chloride, was the only cation which exhibited a significant reduction in cell numbers in high density cultures in spent media. The addition of a one tenth volume (v/v) of 25mM ferric chloride to a CHO cell broth with high viable cell concentration altered the pH to < pH 4 (Table 4.2) and neither calcium chloride nor manganese chloride impacted pH to the same extent. This could be evidence that iron mediates flocculation by adjusting the ionic strength of the solution (by decreasing pH) resulting in the compression of the cells electric double layer thus allowing attractive van der Waal forces between cells to dominate. This theory is supported by the fact that

acidification of the bioreactor contents prior to primary recovery is a recognised strategy by the biologics industry (LeMerdy, 2014)

4.3.2. Calcium Chloride

The flocculating properties of calcium chloride are well known. Calcium chloride has been implicated in the flocculation of brewing yeasts. Initial results indicated that 50mM calcium chloride was insufficient to cause the flocculation of CHO cells. This result is supported by the findings of Westoby et al. (2011), Hägg (2015), Chen et al. (2017) and Burgstaller et al. (2018). Westoby et al. (2011) investigated the effect of pH in conjunction with 10mM calcium chloride on the flocculation of CHO cells. They used a reduction in the turbidity of the CHO broth as a surrogate measurement for cell flocculation. They demonstrated that a pH ≤ 5 was required to obtain a change in turbidity of the CHO broth compared to pH 7. The use of 10mM calcium chloride at pH values above pH 5 did not result in a decrease in broth turbidity compared to pH 7. This observation with respect to the inability of calcium chloride to flocculate cells is replicated by Hägg (2007) who investigated the impact of calcium chloride as a co-coagulant with ferric and aluminium chloride in water treatment and demonstrated that calcium chloride on its own was not as effective as either ferric chloride or aluminium chloride.

Chen et al. (2017) and Burgstaller et al. (2018) used a mixture of calcium and phosphate ions to flocculate CHO cells. These two studies were based on the theory that the combination of calcium and phosphate ions would result in the formation of calcium phosphate which would aid the precipitation of cells out of solution at neutral pH. Figures 4.13a -4.13c inclusive show the effect of 25mM calcium chloride on CHO K1 cells with respect to cell number, cell size and % cellular aggregation. A statistically significant effect, compared to the negative control, was only observed for cell number with 100mM calcium chloride after 10 minutes sedimentation without the need for phosphate (Figure 4.15a). It also appears that there was a dose dependent relationship between cell number and cation concentration as the $8X10^6$ cells/mL cultures demonstrated a larger reduction in total cells/mL for a given concentration of cation

Evidence supporting the theory that calcium is capable of flocculating cells comes from Figures 4.15a, 4.15b, 4.17a and 4.17b. Cells remaining in solution after treatment with

calcium chloride at 100mM showed an increased cell size post incubation with cation while post sedimentation there was no difference in cell size. This correlated with the finding that while there was no significant effect in the total number of cells/mL remaining in solution between the calcium treated cells post cation, there was a significant effect post sedimentation indicating that the larger flocculated cells had settled out as predicted by Stokes law. Cellular aggregation was used as an orthogonal method for measuring cell size. There was a significant increase in cellular aggregation between the negative control and cells treated with 100mM calcium chloride at the post cation stage. This effect disappeared after 10 minutes sedimentation. It's not clear whether this is due cells with a higher level of aggregation settling out or the effect of including EDTA prior to sedimentation. EDTA is known to reduce cell clumping and promote single cell suspensions. Indeed, it's included in trypsin preparations when enzymatically detaching cells from a surface for this very reason.

4.3.3. Manganese Chloride

The use of manganese chloride as a flocculant has been previously reported in yeast cells but not for CHO cells. The use of 25mM manganese chloride resulted in a significant increase in cellular aggregation but this did not translate into a significant effect on the number of cells/mL remaining in solution or the size of those cells. As regards cell numbers and cell size no significant effect on either the total number of cells/mL remaining in solution or cell size post incubation with 50mM manganese chloride was observed. However, post sedimentation a significant effect for both the total number of cells/mL remaining in solution and cell size was observed. It's not clear why a significant effect on cell size post sedimentation was observed. Examination of Figure 4.17b reveals two distinct populations with regards to cell size. Six of the replicates are tightly clustered with an average cell diameter of 12μm while the remaining three range between 14μm - 16μm in diameter. In comparison all nine replicates from the negative control cells are tightly clustered between 11μm and 12μm. While Figure 4.15c shows an increase in cell aggregation for cells treated with 25mM manganese chloride compared to the negative control there isn't a corresponding increase in cell size (Figure 15b). A significant increase in cellular aggregation between the negative control and cells treated with 50mM manganese chloride was observed at the post cation stage. This effect disappeared after 10 minutes sedimentation. It's not clear whether this is due

cells with a higher level of aggregation settling out or the effect of including EDTA prior to sedimentation.

4.3.4. Ferric Chloride

Ferric chloride has a history of being used as a flocculant in the wastewater industry. None of the tested concentrations resulted in a significant effect on cell numbers remaining in solution compared to the negative control for the $8X10^6$ cells/mL culture in fresh media despite a significant effect being observed for the $8X10^6$ cells/mL culture, in spent media, when 5mM ferric chloride was used. This result requires further investigation to see if it's a true result or an outlier. The pattern with regards to cell numbers and cell size was similar to that observed for calcium chloride. There was a significant effect observed for both the total number of cells/mL remaining in suspension and cell size post incubation with ferric chloride. However, post sedimentation, while a significant effect on the total number of cells/mL remaining in solution was observed, no effect on cell size was observed. This indicates that the larger flocs had settled out as predicted by Stokes law. The aggregation data for 25mM ferric chloride contradicts itself. Figure 4.18a showed a significant increase in the level of aggregation among cells remaining in solution post cation incubation compared to the negative control but this effect was reversed post sedimentation (Figure 4.18b). The trend in Figure 4.18b of no significant effect being observed for treated cells was contradicted by Figure 4.15c which shows a significant effect on cellular aggregation compared to the negative control. It's not clear what's behind this difference but a noticeable difference between these two experiments is the inclusion of EDTA in the experiment represented by Figure 4.18b. EDTA is a known chelating agent and could be chelating the iron ions thus causing cellular aggregates to fall apart. Evidence for this theory is shown in Figures 4.23 and 4.24. Figure 4.23 shows greater clearance of cells as indicated by the lack of turbidity in media compared to Figure 4.24. The cell pellet is also larger in volume.

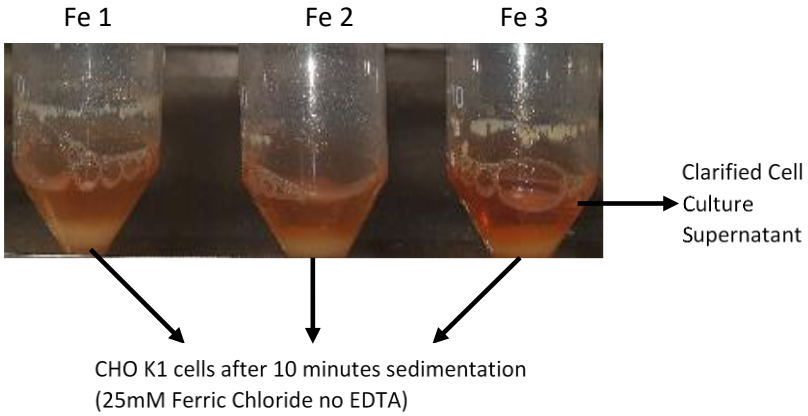


Figure 4.23: **Cells treated with 25mM ferric chloride in the absence of 25mM EDTA post sedimentation. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent) at 2X10^7 cells/mL. A 10X solution of ferric chloride was added to the cells and the cells were incubated at 37^{o}C, 170 rpm followed by cells being left to settle out at room temperature for 10 minutes.**

Figure 4.24: **Cells treated with 25mM ferric chloride and 25mM EDTA post sedimentation. Cells were grown overnight in Balan CD, centrifuged and the cell pellet resuspended in the same media (spent) at 2X10^7 cells/mL. A 10X solution of ferric chloride was added to the cells and the cells were incubated at 37^{o}C, 170 rpm followed by cells being left to settle out at 37^{o}C, 0 rpm for 10 minutes.**

It's worth mentioning that while there is evidence that cell size is related to the osmolality of the media (Alhuthali et al., 2021) this can be excluded as the reason an increase in cell size was observed under the experimental conditions employed in this study. This is because the experiments were performed in batch culture with no pH control. Therefore, any increase in osmotic pressure due to the accumulation of lactate and ammonia should be offset by the corresponding decrease due to the consumption of nutrients such as glucose and amino acids.

4.3.5. Effect of cation addition on antibody yield and quality

To ascertain whether the use of divalent and trivalent cations as flocculating agents impacted on antibody yield a human IgG was spiked into cell broth, incubated with cation, EDTA added, and sedimentation allowed to occur and the cell free antibody containing supernatant recovered. This cell free supernatant was then applied to Protein A/G spin columns and the antibody purified. The antibody yield in the eluates was quantified by ELISA and the purified antibody's quality assessed by gel electrophoresis and SEC-HPLC.

Examination of Figure 4.19c highlights that there was an issue with the purification for one of the replicates of each cation. A technical issue was identified during the purification of replicate three. The manufacturer's instructions for using protein A/G purification columns calls for the clarified cell culture supernatant to be diluted 1 in 2 with binding buffer. This ensures that the pH of the cell free supernatant is close to pH 8 which is the optimum pH for binding of antibodies to protein A/G. However, when this step was performed for replicate three a white precipitate was observed. This precipitate was not observed in the negative control. The pH of the binding buffer was checked and was found to be within specification, so the purification was continued. Antibody was detected in the eluate for each cation and was concentrated at least tenfold using spin columns with a 10kDa molecular weight cut-off prior to performing SDS PAGE and SEC-HPLC analysis.

While a reduction in antibody concentration for each cation was observed at both the post cation and post sedimentation stages compared to the negative control ($FeCl_3$ > $CaCl_2$ > $MnCl_2$) none of the cation treatments resulted in a significant decrease in antibody concentration compared to the negative control. Following purification, the yield of antibody from each of the cations was less than that from the negative control ($MnCl_2$ > $FeCl_3$ > $CaCl_2$). However, there was no significant effect observed for any of the cation treatments and the negative control. It's unlikely that this lack of an effect is due to the binding capacity of the Protein A/G column being exceeded. According to the manufacturer, the binding capacity of the Protein A/G columns used in this experiment was approximately 7g. The amount of antibody loaded on the column varied from 3.7g ($FeCl_3$) to 6.0g (Negative Control).

The eluates from each cation were analysed on a 4-12% gradient reducing and non-reducing SDS gels (Figures 4.20 and 4.21) and visually compared. Crucially there's no evidence of product dissociation which is a known risk associated with lack of agitation of the cell broth during harvest and purification operations.

SEC-HPLC was used to calculate the molecular weight of the antibody monomer peak and quantify the amount of aggregation and fragmentation in each eluate. The molecular weight of each peak in the chromatogram was calculated by reference to the retention time of standards of known molecular weight. Retention time was plotted against the log molecular weight of four standards and a calibration curve constructed. This curve had a linear regression coefficient of 0.98. The average calculated molecular weight of the monomer peak was between 114kDa and 116kDa for the negative control and the three cations. This is lower than the theoretical molecular weight of an IgG antibody (150kDa) and was attributed to measurement uncertainty rather than evidence of antibody degradation. This theory is supported by the fact that the calculated molecular weight for one of the standards (gammaglobulin, 150kDa, 8.95 minutes) was 121kDa according to the calibration curve.

Treatment of cells with 25mM ferric chloride showed a significant decrease in % monomer and a corresponding significant increase in % fragments and aggregates compared to the negative control. This supports the non-reduced SDS PAGE analysis where additional high molecular weight bands were observed. It also supports the idea that the reduction in antibody yield associated with 25mM ferric chloride is due to iron binding to the antibody and forming protein aggregates.

Iron (II) ions can react with hydrogen peroxide to generate hydroxy radicals (HO●) via a reaction known as the Fenton reaction. These hydroxy radicals have can subsequently interact with cells, DNA and proteins resulting in oxidative damage to all amino acid residues but primarily sulphur containing amino acids such as methionine, tryptophan and cysteine. From a bioprocess perspective the impact of such changes includes functional alterations, loss of structural integrity, protein aggregation and increased product heterogeneity, all of which will reduce the quality and yield of purified antibody.

Although iron (III) was the form of iron ion used in this study, Walling and Weil (1974) demonstrated that this ion can undergo Fenton like reactions to generate hydroperoxyl

radicals (HO_2●) and iron(III) ions, albeit at a rate four orders of magnitude lower than that of the Fenton reaction. Perhaps more worryingly is the report by Gutteridge et al. (1990) that a complex of iron (III)-EDTA can react with hydrogen peroxide to generate hydroxy radicals (HO●).

An investigation into the potential of iron(III) and EDTA to cause oxidative damage to recombinant proteins would have to be conducted if the approach explored in this study is to be applicable to the biotech industry. Isoelectric focusing techniques such as imaged capillary isoelectric focusing (icIEF) would be the standard approach in the biologics industry for studying if changes to a recombinant proteins charge had occurred. However, mass spectrometry is an orthogonal approach which would have the added advantage of identifying the exact peptide which was modified, and would be proposed for future studies.

4.4. Conclusions

The antibody yield achieved in this study doesn't support the approach examined in this chapter as a valid method for cell clarification at an industrial scale. Currently there's an industry wide focus on the cost of goods. Cost of goods measures how expensive it is to produce a g of recombinant protein. Industrywide there's a goal of reducing the cost of goods to below UD$10/g recombinant protein. While many factors influence the cost of goods figure, running a process in which half of the recombinant protein is lost during harvest is not conducive to meeting this UD$10/g target. Percentage recovery of the spiked in antibody ranged from 44.7% with manganese chloride to 53.4% for calcium chloride.

Calcium chloride at 100mM was capable of flocculating CHO cells at $2X10^7$cells/mL as measured by the reduction in the total number of cells/mL after sedimentation.

Manganese chloride at 50mM was capable of flocculating CHO cells at $2X10^7$cells/mL as measured by the reduction in the total number of cells/mL after sedimentation.

Ferric chloride at 25mM was capable of flocculating CHO cells at $2X10^7$cells/mL without the need for sedimentation as measured by the reduction in the total number of cells/mL post incubation with cation. Allowing cells which had been incubated with 25mM ferric chloride to sediment for 10 minutes achieved total clarification of the cell broth on a visual basis. Adding 25mM EDTA prior to sedimentation reversed this finding.

The product quality assays don't offer any evidence that the compounds chosen as flocculating agents in this study will result in fundamental changes in antibody structure such as dissociation, cleavage of disulphide bonds or metal catalysed antibody degradation. However, treatment with 25mM ferric chloride did result in increased antibody aggregation.

The fact that ferric chloride had a significant effect on cell numbers at the post cation stage combined with the fact that it significantly lowers the media pH supported the theory that it acted by compression of the cells dielectric layer.

Conversely, the fact that CHO K1 cells treated with calcium or manganese need to be left to settle before a significant reduction in cell number was achieved supported the theory that charge bridging between either the calcium or manganese ion and the CHO K1 cell was the mechanism responsible for the flocculation of CHO K1 cells treated with calcium chloride or manganese chloride.

The finding that fresh media caused a greater reduction in cell numbers than spent media for a given concentration of cation was unexpected. While implementing a media change strategy, as a means of promoting flocculation, is not viable in a manufacturing process employing stirred tank reactors, it may have an application with perfusion-based manufacturing processes. This is because it would be a relatively simple operation to change the media reservoir on a perfusion reactor prior to initiating cell harvest operations. Pilot experiments would need to be run to see if the finding that fresh media promotes a greater reduction in cells numbers is unique to Balan CD or hold for all media.

CHAPTER 5

A proteomic comparison of suspension adapted CHO K1 cells grown as attached and suspension cultures in the presence and absence of FBS

5.1. Introduction

In the previous chapters the challenges associated with the clarification of high density cell lines in industrial settings were discussed and two potential solutions which employed flocculation of the bioreactor contents prior to performing clarification steps were investigated. Chapter four investigated the feasibility of using divalent and trivalent cations as flocculating agents while chapter three investigated the cloning and attempted expression of four eukaryotic proteins exogenous to CHO cells as a means of developing a flocculating CHO cell line. However, none of the proteins investigated appeared to be capable of driving flocculation in CHO cells. Therefore, we theorised that a proteomic based approach may identify proteins endogenous to CHO cells which could be used to drive cellular flocculation.

The separation and identification of individual proteins from complex mixtures has been revolutionised by the development of highly sensitive mass spectrometers, e.g., Orbitrap, which can be coupled to HPLC systems (LC-MS). These types of instruments when combined with label free quantification method (Section 1.4.4) s and data analysis techniques such as Peptide Set Enrichment Analysis (PSEA) has allowed for the identification and quantification of changes in the expression level of individual proteins between two or more samples. PSEA is a modification of Gene Set Enrichment Analysis (GSEA). GSEA, a technique for the identification of genes which have changed their expression level between two conditions (control and test), has been applied to the analysis of a proteomic dataset by Isserlin et al. (2010). However, the greater variability associated with measuring protein abundance from a mass spectrometry experiment, compared to measuring gene abundance in a microarray experiment, meant that GSEA was not the ideal approach to the analysis of proteomic data. Peptide set enrichment analysis (PSEA), an adaption of GSEA, which overcame the challenges associated with the application of GSEA to the analysis of proteomic data was reported by Cha et al. (2010) and Grigorieva et al. (2020). Instead of looking at changes in gene expression PSEA looks for consistent correlations across a group of proteins. In this way it's possible to identify a significant biological effect that is smaller in magnitude (per individual protein) than that which could be identified by examining single proteins in isolation.

This type of experimental approach, when allied with modern bioinformatic databases such as DAVID (Dennis et al., 2003), Gene Ontology (Ashburner et al., (2000) and KEGG pathways Kanehisa et al., 2014 and Kanehisa and Goto, 2000) has revolutionised the study of protein expression and allows an in depth understanding of the biological themes contained within a proteomic differential expression study to be acquired.

One of the reasons why CHO cells became the predominant cell line used in the biopharmaceutical industry was the development of CHO cell lines capable of suspension growth in serum free chemically defined media. This has allowed the development of high titre, economically viable processes at scales of up to 20,000L in fed batch stirred tank reactors. Various "omic" studies have been reported in the literature describing CHO cells in general and CHO cells growing in suspension in particular.

- Nissom et al. (2006) compared CHO cells lines that differed in their ability to produce a *dhfr*-GFP fusion protein.
- Meleady et al. (2011) compared two antibody producing CHO cell lines that differed in their specific productivity (Qp) throughout a ten-day fed batch culture.
- Baycin-Hizal et al. (2012) described the CHO proteome, secretome and glycoproteome and linked them to the CHO genome, thus increasing the accuracy of protein identification.
- Heffner et al. (2017) extended the approach of Baycin-Hizal et al. (2012) by comparing the CHO K1 proteome to the proteome of specific hamster tissues.
- Henry et al. (2017) used phosphoproteomics to profile CHO cells grown at two different temperatures.
- Xu et al. (2017) compared the proteome of CHO K1, CHO S and a *dhfr*$^{-}$ CHO cell line.
- Kaushik et al. (2018) used phosphoproteomics to characterise differences in the phosphoproteome of CHO DP12 cells between the exponential, stationary, and decline growth phases.
- Hausmann et al. (2018) compared the CHO proteome before and after selection with methotrexate.

- Valente et al. (2018) reviewed the application of proteomic approaches to CHO HCP characterisation.
- Bertrand et al. (2019) used an MS proteomic approach to monitor the cell culture performance for a CHO cell line expressing a monoclonal antibody in a perfusion bioreactor.
- Chevallier et al. (2020) compared differences in the ability of cells to metabolise Glutathione (GSH) and the relationship between intracellular GSH and recombinant product titre.
- Heffner et al. *(*2020) expanded on their 2017 study.
- Klingler et al. (2021) used an MS/MS approach to identify mechanisms involved in cell aggregation.
- Bryan et al. (2021) used a proteomic approach to investigate the molecular basis of growth-related phenotypes in CHO cell lines.
- Strasser et al. (2021) investigated the effect of bioprocessing conditions on productivity and product quality in a CHO cell line producing an IgG_1.
- Park et al. (2021) used proteomics to investigate the differences in the CHO proteome between cells which produced antibodies susceptible to disulphide bond reduction and those which weren't.

The hypothesis under investigation in this chapter, was that the biological processes underpinning cellular attachment had been disrupted during the adaptation of CHO K1 cells from attached growth to suspension growth, and that an in-depth understanding of the proteomic changes associated with these processes may reveal proteins which could be engineered to promote cellular flocculation immediately prior to harvest. Therefore, suspension adapted CHO K1 cells were grown as attached and suspension cultures and their proteomes compared in a LC-MS label free differential expression experiment (Section 1.4.4). This approach was based on the observation that while CHO cell lines capable of growing, as single cells in suspension culture, with high titres and viability at harvest have been developed, the initial CHO cell line isolated by T. Puck in 1957 (Puck et al., 1958) was established as an attached cell line. The adaptation of CHO cells from attached to suspension growth involved both phenotypic and genotypic changes. The phenotypic change of interest to this study was that CHO cells no longer grew as elongated cells attached to a surface but instead grew as individual rounded

cells in suspension. However, given the observation made above, regarding the changes in morphology observed when comparing attached CHO K1 cells with suspension adapted CHO K1 cells, it seems logical to anticipate that the expression of genes involved in cellular attachment have been disrupted. These genes maybe candidates for driving cellular flocculation and being endogenous to CHO shouldn't suffer from the challenges associated with expressing exogenous genes as experienced in chapter three.

5.2. Experimental Design and Sample Preparation

In this chapter a suspension adapted CHO K1 cell line was grown as both suspension and attached cultures in the presence and absence of serum. Whole cell lysates were prepared and their proteomic profiles analysed in a label-free quantitative differential expression experiment. Cellular processes involve multiple proteins working synergistically. Therefore, it was considered likely that simultaneous small changes in the expression of multiple proteins could result in a larger effect than a large change in the expression of a single protein.

The PSEA method described by Ulgen et al. (2019) was used to perform PSEA on lists of differentially expressed proteins resulting from comparing attached cells grown in 5% FBS with suspension cells grown in 5% FBS and also comparing attached cells grown in the absence of FBS with suspension cells grown in the absence of FBS. This enabled the identification of statistically enriched KEGG pathways, the DE proteins implicated in those pathways and the upregulated or downregulated status of each DE protein for both comparisons. The genePieList R package was used to categorise the enriched biological processes identified from each comparison (Figures 5.8 and 5.9).

It's expected that this approach will provide an insight into the proteomic changes underpinning this adaptation process.

To identify proteins which differed in their expression levels between attached and suspension cultures a comparative analysis was performed using a label-free quantitative differential expression LC-MS approach. The approach taken utilised four biological replicates of each condition and is outlined in Figure 5.1.

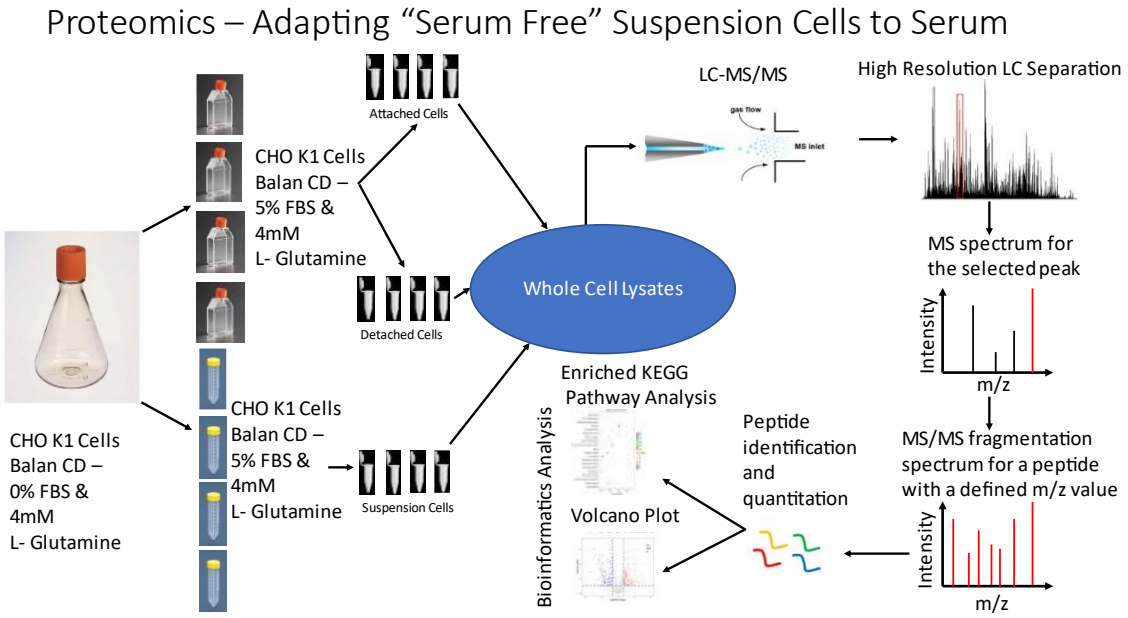


Figure 5.1: **Workflow followed when preparing samples grown in 5% FBS for LC-MS analysis. The same approach was followed for cells grown in 0% FBS.**

A non-producing suspension adapted CHO K1 cell line was chosen for this experiment. These cells were grown under the following two conditions:

1. CHO K1 cells pre-adapted to suspension were grown as attached cells in T75 flasks in a chemically defined media (Balan CD growth medium A) in the presence (5%) and absence (0%) of FBS, seeded at $2.5X10^5$ cells/flask. A T75 flask per FBS concentration was split on day 4 and used to seed four T175 flasks with vented lids for each FBS concentration with $1X10^6$ cells in total per flask. These were grown until they were approximately 80% confluent. Upon harvest these cells were further sub divided into two sub populations, attached and detached cells. Attached cells is a self-explanatory term. These cells remained attached to the surface of the T175 flask and were removed by scrapping. The phrase "detached cells" described cells which spontaneously detached from the surface of the flask but were shown to be viable cells floating in the media. These cells were recovered by decanting the media. All cell populations were pelleted by centrifugation, washed with ice cold PBS and the pellets stored at -80^{o}C until required.

2. Suspension CHO K1 cells were cultivated in 250mL Corning® shake flasks (50mL working volume) using Balan CD growth medium A in a Kühner orbital shaker at 37°C, 5% CO_2, 80% RH and 170rpm. For proteomic sample preparation four TPP tubes per FBS concentration were seeded with 5mL of cells at $2X10^5$ cells/mL ($1X10^6$ cells in total per tube). These were grown for 3 days, the cells pelleted, washed with ice cold PBS and the pellets stored at -80°C until required.

The sample preparation for any proteomics experiment will comprise the following steps:

1. Lysis of cell pellets to release proteins.
2. Quantification/normalisation of released proteins between samples.
3. Denaturation, alkylation and digestion of the proteins to produce peptides.
4. Resuspension of the peptides in a mass spec compatible buffer.

In this project these steps were achieved utilising the FASP protocol developed by Coleman et al. (2017) (Section 2.6.3.7).

To reduce variability and retention time drift, all timepoints corresponding to each cell line were prepared and analysed on the mass spectrophotometer in a single analysis. Following peptide separation, on a C18 column, and MS analysis, using a Thermo Orbitrap Fusion Tribrid Mass Spectrometer, raw MS data files were imported into the Progenesis QI for Proteomics software and automatic run alignment was performed to minimise variability between samples and replicates. Kelly (2016) described two strategies to compare data from label free experiments. These were "Global Analysis" and "Two-Sample Analysis''. Two sample analysis compares two cell lines or experimental groups against each other. Global analysis compares all samples against each other as a single group. It allows for differences in the expression levels of individual proteins to be tracked across multiple cell lines/experimental groups. The "Two-Sample Analysis Approach'' was adopted in this study.

5.2.1. Quantitative label free LC-MS/MS Analysis - Identification of DE peptides.

Raw MS data files were imported into the Progenesis QI for Proteomics (Nonlinear Dynamics) software and automatic run alignment was performed to minimise variability between replicates. Following alignment, samples were assigned to experimental groups and identified features were filtered based on an ANOVA p value of ≤ 0.05 between experimental groups. This ensures that only statistically significant differentially expressed peptides are identified and selected for further analysis. A mascot generic file (.mgf) of significantly differentially expressed peptides was generated for each comparison and this file was exported to Proteome Discover for protein identification.

5.2.2. Quantitative label free LC-MS/MS Analysis – Protein assignment.

The SEQUEST (HT) algorithm was used to identify individual peptide sequences, and the proteins they were derived from, by searching against a proteogenomic CHO database (Section 2.6.5) (Li et al., 2018). The resulting protein lists were imported back into Progenesis QI for Proteomics and only proteins with an ANOVA p value of ≤ 0.05, a relative minimum fold change ≥ 1.2 between experimental groups and at least two independent peptides contributing to the identification were identified as differentially expressed. Finally, principal component analysis and heatmaps based on the Euclidean distance were used to assess clustering of replicate samples. (Section 2.6.8)

All subsequent data manipulation, protein naming, and statistical analysis was performed in the R programming environment (Section 2.6.6).

5.3. Results

The experimental design followed gave rise to several distinct experimental groups:

1. Suspension cells grown in Balan CD supplemented with 5% FBS (S5)
2. Suspension cells grown in Balan CD in the absence of FBS (S0)
3. Attached cells grown in Balan CD supplemented with 5% FBS (A5)
4. Attached cells grown in Balan CD in the absence of FBS (A0)

5. Detached cells grown in Balan CD supplemented with 5% FBS (F5)
6. Detached cells grown in Balan CD in the absence of FBS (F0)

Detached cells are those cells which were detached from the surface of the T175 flask but were verified as being viable using the Guava ViaCount^{TM} assay for viability. This definition is provided for clarity only. These cells were analysed by MS/MS-HPLC but weren't include in any differential protein analysis or comparisons as discussed below. Hence the viability data hasn't been included in this thesis.

Consideration of the experimental groups chosen for analysis resulted in the decision to exclude the detached cells from the analysis. These cells were observed as a sub population of the attached cells. Initially it was considered that these cells represented cells which had spontaneously detached from the surface of the cell culture flask. However, given that the cells used to generate the "attached cells" experimental group were cells preadapted to suspension growth, it's more likely that these detached cells never attached in the first place. Accordingly detached cells have not been included in any analysis described in this chapter.

It was decided to compare cultures grown with which differed based on culture format while keeping the serum concentration constant. The DE protein list for each experimental group was subjected to the following analysis:

a) Heatmaps and principal component analysis (PCA) were used to visualise the clustering pattern of replicate attached and suspension cultures. This is discussed in greater detail in Sections 2.6.7 and 2.6.8.
b) Analysis of whether all DE proteins detected in each experimental group were upregulated or downregulated in suspension culture compared to attached culture using a volcano plot.
c) Peptide set enrichment analysis of the data using the pathfindR R package to identify statistically enriched KEGG pathways ($p \leq 0.05$) and whether the proteins identified from those pathways were upregulated or downregulated.
d) Identification of enriched Gene Ontology (GO) terms ($p \leq 0.05$) which describe the DE proteins identified by PSEA using the enrichGO function from the Annotation Hub R package.

e) Categorisation of enriched GO terms with respect to the cellular biological process they were involved in using the geneListPie R package.

5.3.1. Assessment of the clustering pattern of replicates

One of the features of the Progenesis QI for Proteomics software was that it utilised principal component analysis (PCA) to analysis the grouping of replicate injections (Section 2.6.8). PCA is a multivariate analysis that aims to explain the variance of the data by transforming it to a new coordinate system (each coordinate called a principal component) usually after a normalisation step. This allows the maximum amount of information in a dataset to be described using the least possible number of variables (principal components). It's standard practice for principal component 1 to describe the largest amount of variation in the data and sufficient components are assigned until all the variation is described. The PCA model describing the variation in the data is considered a good fit if most of the variation in the dataset is described by principal components 1 and 2. This simpler picture can help assess the quality of the data (i.e., replicate samples should cluster together) and can help in the interpretation of the data by matching variables with components. It has as its aim increasing ease of interpretation while minimising information loss. In an ideal scenario, all the variation in a dataset would be described by one variable (principal component 1), but this rarely (if ever) occurs. The specific PCA model used by the Progenesis QI software used the abundance level of each individual DE protein to determine the variation in the data and therefore experimental groups and replicates can be separated based on differences in the abundance of each protein. This aids the identification of outliers.

PCA analysis of suspension adapted CHO K1 cells grown as attached and suspension cells in media containing 5% FBS supports the theory that the change in culture format and or the presence of serum in the media induced changes on proteomic level as two distinct clusters corresponding to the different culture formats can be observed in Figure 5.2. The replicate suspension cultures cluster together and the variability within them is described by principal component one. However, only three of the replicate attached cultures cluster together. The variability for the attached cultures is explained by both principal component one and two. This lack of clustering among the attached cultures is

visual evidence of proteomic changes that these cultures have undergone as they adapted to growing attached to a surface in the presence of serum. Figure 5.3 presents a similar picture whereby the suspension cultures cluster tightly together while the attached cultures are more spread out. There is a difference between the attached and suspension cells with regards to which principal components described the variability within the two experimental groups. Principal component one describes the variability within the suspension cells while principal component two describes the variability with the attached cultures.

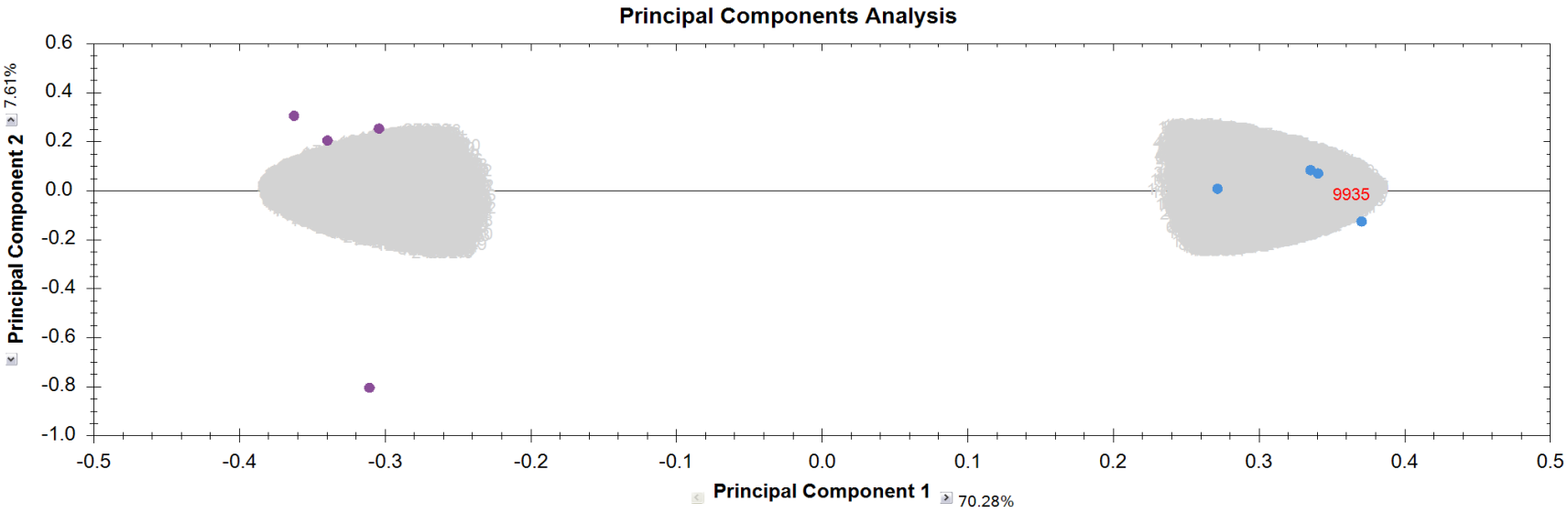


Figure 5.2: **PCA plot showing the DE proteins associated with all 4 replicate injections for the CHO K1 cells grown as attached and suspension cultures in media containing 5% FBS. Blue dots are the attached cultures while purple dots are suspension cultures.**

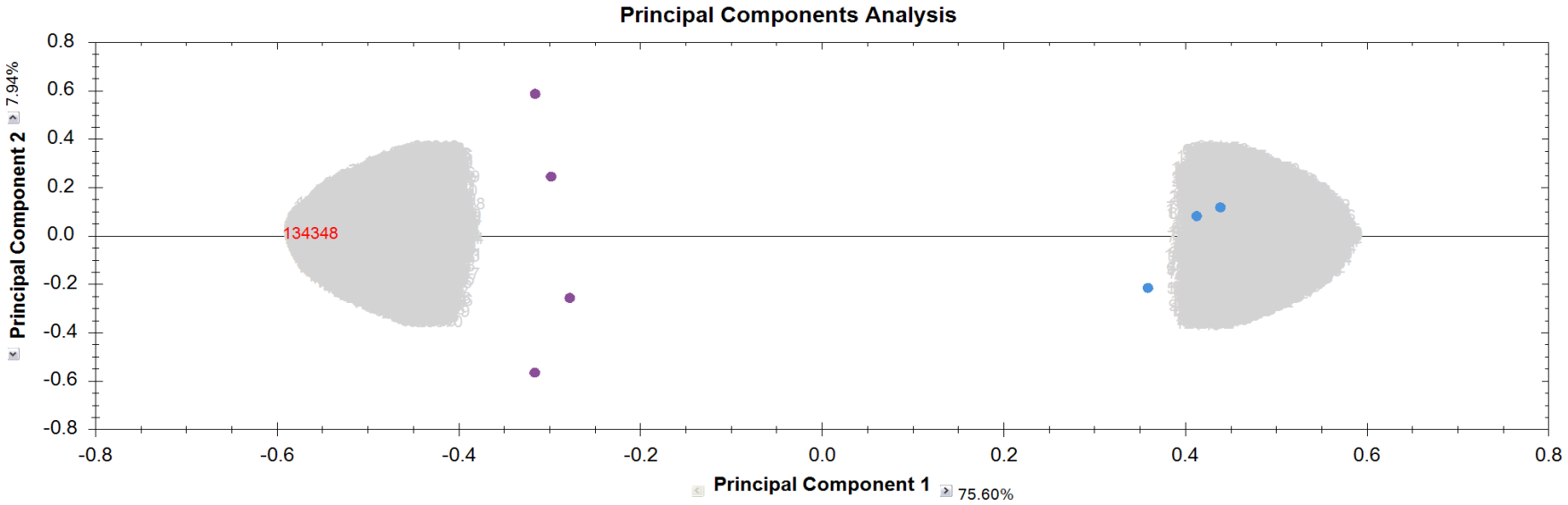


Figure 5.3: **PCA plot showing the DE proteins associated with all 4 replicate injections for the CHO K1 cells grown as attached and suspension cultures in media without FBS. Blue dots are the attached cultures while purple dots are suspension cultures.**

The clustering of the two experimental groups was further examined using a heatmap based on the Euclidean distance. All four replicates from each experimental group (attached cells 5% FBS, suspension cells 5% FBS, attached cells 0% FBS and suspension cells 0% FBS) were analysed (Figure 5.4 and 5.5).

Examination of Figure 5.4 revealed three clusters as follows:

1. Cluster one – S0a, S0d and A0a.
2. Cluster two – S0c and S0b.
3. Cluster three – A0b, A0c and A0d.

Based on the dendrogram levels clusters one and two are the most closely related of the three clusters. This suggests that A0b, A0c and A0d have started expressing changes in its proteome relative to the S0 cultures reflecting the fact that Ab, A0c and A0d have potentially started adapting to growth as attached cells while A0a hasn't and retains a proteome more associated with cells growing in suspension.

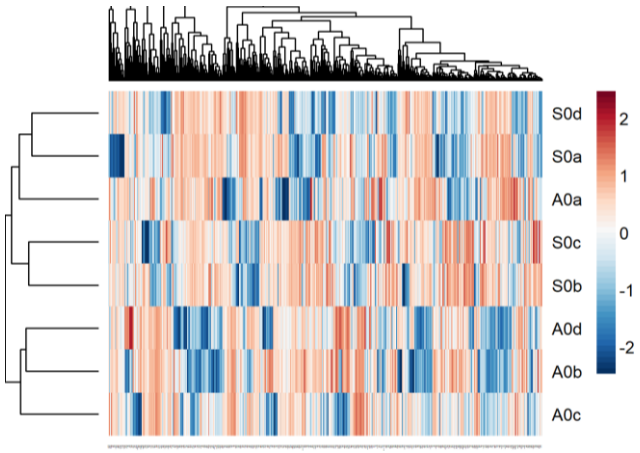


Figure 5.4: **Heatmap based on Euclidean distance. All four replicate injections from the attached cultures + FBS (A0) and suspension cultures – FBS (S0) are presented. The letters a, b, c, and d refer to individual replicates.**

Two clusters can be identified in Figure 5.5. :

1. Cluster one comprises of the four attached replicate cultures and one suspension culture S5c.

2. Cluster two comprises of the remaining three suspension replicate cultures.

This differs from the pattern in Figure 5.4. The inclusion of serum in the culture media has resulted in the four attached cultures clustering together along with one of the suspension cultures.

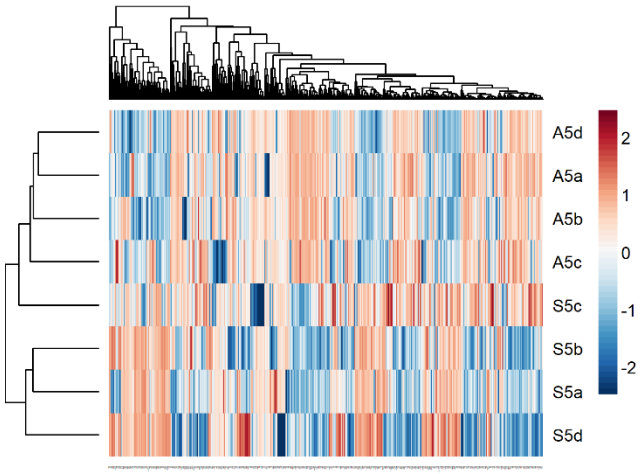


Figure 5.5: **Heatmap based on Euclidean distance. All four replicate injections from the attached cultures + FBS (A5) and suspension cultures – FBS (S5) are presented. The letters a, b, c, and d refer to individual replicates.**

It is possible that the clustering pattern observed in in Figures 5.4 and 5.5 is due a lack of reproducibility during cell recovery from the flask, cell lysis and sample preparation. It should also be noted that although biological replication (multiple flasks were sampled) was considered during the experimental design, technical replication (multiple injections from the same MS sample vial) wasn't. Mechanical scraping was the method used to remove cells from the surface of the flask. While an equal number of cells was not recovered from each flask, cell numbers were equalised prior to cell lysis so that an equal number of cells was lysed for each replicate. Equally protein concentration was equalised across replicates prior to sample preparation. Obviously, if the total number of proteins detected by the mass spec differs between replicates of the same sample or if a particular protein is not detected in all replicates of a sample this will increase variability and could potentially lead to a protein being identified as differentially expressed in error. To investigate this, the total number of proteins detected in each replicate and the

number of proteins common to all four replicates of each experimental group was quantified using Proteome Discoverer. These numbers are displayed in Table 5.1.

Experimental Group	Replicate	Total No. of Proteins	Average No. of Proteins	No. of Proteins Common to each replicate	No. of DE Proteins
Attached (5% FBS)	A5a	3814	3753	2858	1118 FC = 1.2
	A5b	3720			
	A5c	3732			
	A5d	3745			
Suspension (5% FBS)	S5a	3733	3746	2828	
	S5b	3619			
	S5c	3868			
	S5d	3765			
Attached (0% FBS)	A0a	3780	3401	2436	1158 FC = 1.2
	A0b	3297			
	A0c	3410			
	A0d	3116			
Suspension (0% FBS)	S0a	3598	3706	2735	
	S0b	3786			
	S0c	3670			
	S0d	3769			

Table 5.1: **Total numbers of proteins detected in all replicates. MS raw data files were imported directly into Proteome Discoverer and proteins were detected as per Section 2.6.5. The number of common proteins column displays the number of proteins detected in all four replicates.**

In excess of three thousand proteins were detected in each of the replicates in both the A5vS5 and A0vS0 experimental groups. However, not all of these proteins were detected in all replicates. For the A5vS5 experimental group 2858 and 2828 proteins were detected in all attached and suspension replicates respectively. The numbers for the A0vS0 experimental group were 2436 (attached) and 2735 (suspension) respectively. In order to assess the impact of this, the proteins which were identified as common to the experimental groups were compared to the DE proteins identified as playing a role in the KEGG pathways identified in section 5.3.3. This enabled the identification of any proteins

present in these pathways which were not common to all replicates. These are listed in Table 5.2.

KEGG Path	HGNC Symbol	Direction	Experimental Group
HIF-1 signalling pathway	PFKP	Up regulated	A0vS0
	PLCG1		A5vS5
Retrograde endocannabinoid	PTGS2, NDUFB8	Down regulated	A0vS0
	PTGS2		A5vS5
Glucagon signalling pathway	PPP3R1, PYGM, PFKP	Up regulated	A0vS0
	PPP3R1		A5vS5
VEGF signalling pathway	PTGS2	Down regulated	A0vS0 and A5vS5
	PPP3R1	Up regulated	A0vS0 and A5vS5
	PLCG1		A5vS5
Estrogen signalling pathway	KRT14, KRT17, MMP9	Down regulated	A0vS0 and A5vS5
	HSPA2, GRB2	Up regulated	A5vS5
Insulin signalling pathway	PYGM	Up regulated	A0vS0
	GRB2		A5vS5
Oxytocin signalling pathway	PTGS2	Down regulated	A0vS0 and A5vS5
	PPP3R1	Up regulated	A0vS0 and A5vS5
	PPP1R12A		A5vS5
cGMP-PKG signalling	PPP3R1, PPP1R12A	Up regulated	A5vS5
Thyroid hormone	PLCG1	Up regulated	A5vS5
ErbB signalling	PLCG1, PAK2, GRB2	Up regulated	A5vS5
AMPK signalling pathway	PFKP	Up regulated	A0vS0
	HMGCR	Downregulated	
Protein names have not been used for space reasons. Instead, the proteins names associated with all HGNC symbols mentioned in this chapter have been listed in Table 5.9 at the end of this chapter.			

Table 5.2: **DE proteins detected in enriched KEGG paths which were not detected all eight replicates of each condition.**

Fifteen unique proteins were identified as being part of the KEGG pathways identified in Section 5.3.3 which were not detected in all replicates of each experimental group.

5.3.2. Identification of differentially expressed proteins in suspension adapted CHO K1 cells grown as both attached and suspension cultures in the presence and absence of serum.

One thousand one hundred and eighteen differentially expressed proteins were identified when comparing the attached and suspension CHO K1 cultures grown in media containing 5% FBS. The comparison of attached and suspension CHO K1 cultures grown in media containing without FBS identified one thousand one hundred and fifty-eight differentially expressed proteins. The top ten upregulated and downregulated DE proteins based on both fold (FC) and p value for the A5vS5 and A0vS0 experimental groups are presented in Table 5.3 and Table 5.4 respectively.

HGNC Symbol	Comparison	Status	p value	log_2FC	FC	Present in all Replicates	Comment
DE Proteins ranked by p value							
RRP1B	A5vS5	Upregulated	5.57E-08	1.38	2.61	Yes	None
HMGCR			3.57E-07	2.02	4.06	No	Not detected in S5 replicates A and D
SQLE			7.95E-07	1.73	3.32	Yes	None
NONO			1.79E-06	2.32	4.98	Yes	None
AKAP1			1.89E-06	1.34	2.53	Yes	None
STT3B			2.21E-06	1.53	2.89	Yes	None
PTPN1			2.31E-06	1.61	3.06	Yes	None
CNN3			2.41E-06	1.69	3.23	Yes	None
NOLC1			3.23E-06	1.80	3.48	Yes	None
TMTC3			4.56E-06	2.15	4.43	No	Not detected in S5 replicate A
DE Proteins ranked by p value							
THBD	A5vS5	Downregulated	7.30E-07	-1.84	0.28	Yes	None
PTTG1IP			1.31E-06	-0.81	0.57	No	Not detected in A5 replicate B
VIM			2.59E-06	-0.74	0.60	Yes	None
LUC7L2			3.57E-06	-0.70	0.61	Yes	None
SLC2A1			4.09E-06	-1.13	0.46	Yes	None
P4HA1			6.05E-06	-1.04	0.49	Yes	None
SLC6A6			6.92E-06	-0.84	0.56	Yes	None

HGNC Symbol	Comparison	Status	p value	log_2FC	FC	Present in all Replicates	Comment
PBXIP1	A5vS5	Downregulated	7.75E-06	-0.95	0.52	Yes	None
RHOG			9.00E-06	-0.89	0.54	Yes	None
LGALS1			9.54E-06	-1.27	0.41	Yes	None
DE Proteins ranked by log_2 FC value							
UTP20	A5vS5	Upregulated	4.39E-04	3.91	15.01	No	Not detected in any S5 replicate
LOXL1			9.20E-05	3.21	9.26	No	Not detected in any A5 replicate
EIF4H			1.80E-04	2.96	7.76	No	Not detected in A5 replicates C and D
PLIN2			9.61E-06	2.75	6.72	No	Not detected in S5 replicates A and B
SLC25A1			5.66E-06	2.33	5.03	Yes	None
NONO			1.79E-06	2.32	4.98	Yes	None
GPAT3			1.60E-04	2.26	4.79	No	Not found in A5 replicates A, C, D or S5 replicates A, B and C
TMTC3			4.56E-06	2.15	4.43	No	Not detected in S5 replicate A
CRNKL1			2.04E-03	2.14	4.42	No	Not detected in A5 replicates A, D or S5 B and D
SCD1			2.54E-04	2.13	4.39	Yes	None
DE Proteins ranked by log_2 FC value							
NEU2	A5vS5	Downregulated	6.75E-05	-3.31	0.10	No	Not detected in A5 replicates B and D
TSN			2.27E-04	-2.36	0.20	Yes	None
KIF20B			6.56E-04	-2.28	0.21	No	Not detected in A5 replicates B and C

HGNC Symbol	Comparison	Status	p value	log_2FC	FC	Present in all Replicates	Comment
MCM5			5.97E-05	-2.22	0.22	Yes	None
SDF2			3.62E-04	-2.04	0.24	Yes	None
KDM2A			3.59E-05	-2.04	0.24	No	Not detected in A5 replicate B
CLIP1			3.55E-03	-2.01	0.25	No	Not detected in A5 replicates B and D
UHRF1			1.39E-03	-1.92	0.26	Yes	None
RBM3			1.63E-04	-1.92	0.27	No	Not detected in A5 replicate B
TMEM132A			1.86E-03	-1.88	0.27	Yes	None
Protein names have not been used for space reasons. Instead, the proteins names associated with all HGNC symbols mentioned in this chapter have been listed in Table 5.9 at the end of this chapter.							

Table 5.3: **Top 10 upregulated and downregulated DE proteins for A5 v S5 comparisons ranked by pvalue and Log_2FC value from Figure 5.6.**

HGNC Symbol	Comparison	Status	Pvalue	log_2FC	FC	Present in all replicates	Comments
DE Proteins ranked by p value							
SNX2	A0 v S0	Upregulated	1.89E-07	0.93	1.90	Yes	None
LPL			1.90E-07	2.62	6.15	Yes	None
YARS			2.38E-07	1.74	3.33	Yes	None
LTA4H			6.40E-07	1.09	2.13	Yes	None
TALDO1			1.27E-06	1.63	3.10	Yes	None
MDH1			1.31E-06	1.05	2.07	Yes	None
VPS29			1.38E-06	0.83	1.78	Yes	None
XPNPEP1			1.64E-06	1.94	3.84	Yes	None
PTPA			2.29E-06	**1.46**	2.75	Yes	None
LGMN			3.22E-06	1.57	2.96	Yes	None
DE Proteins ranked by p value							
NCOA5	A0vS0	Downregulated	3.31E-07	-1.18	0.44	No	Not detected in A0 replicate A
THBD			6.86E-06	-2.24	0.21	Yes	None
HSD17B7			7.95E-06	-0.64	0.64	Yes	None
EDEM3			1.53E-05	-0.67	0.63	Yes	None
TMED5			3.23E-05	-1.02	0.49	Yes	None
NDUFAF1			3.56E-05	-0.85	0.55	Yes	None

HGNC Symbol	Comparison	Status	Pvalue	log_2FC	FC	Present in all replicates	Comments
ENDOD1	A0vS0	Downregulated	3.91E-05	-0.75	0.59	Yes	None
HMOX1			4.94E-05	-1.93	0.26	Yes	None
COX6C2			5.10E-05	-1.08	0.47	Yes	None
AXL			5.73E-05	-1.48	0.36	Yes	None
DE Proteins ranked by log_2 FC value							
EFEMP1	A0 v S0	Upregulated	6.13E-05	3.74	13.34	Yes	None
ACSS2			1.05E-04	3.28	9.71	Yes	None
NT5DC1			1.56E-04	2.86	7.25	No	Not detected in S0b
LPL			1.90E-07	2.62	6.15	Yes	None
RNMT			8.86E-04	2.56	5.89	No	Not detected in A0a
ALAD			4.00E-04	2.39	5.25	No	Not detected in A0b, A0d
FAM107B			4.06E-03	2.38	5.22	Yes	None
CTPS2			4.47E-06	2.36	5.12	Yes	None
GSTM7			4.89E-05	2.33	5.03	Yes	None
SEMA3B			8.94E-06	2.31	4.95	Yes	None

HGNC Symbol	Comparison	Status	Pvalue	log_2FC	FC	Present in all replicates	Comments
DE Proteins ranked by log_2 FC value							
HMGCR	A0vS0	Downregulated	9.45E-05	-2.59	0.17	Yes	None
JUP			1.43E-03	-2.54	0.17	Yes	None
THBD			6.86E-06	-2.24	0.21	No	Not detected in S0b
FAR1			1.17E-03	-2.21	0.22	Yes	None
TMEM129			5.16E-04	-2.17	0.22	No	Not detected in A0a
MCL1			2.39E-03	-2.14	0.23	No	Not detected in A0b, A0d
PTGS2			1.10E-02	-2.09	0.24	Yes	None
HMOX1			4.94E-05	-1.93	0.26	Yes	None
PLAUR			8.33E-05	-1.93	0.26	Yes	None
NCAM1			3.20E-04	-1.82	0.28	Yes	None
Protein names have not been used for space reasons. Instead, the proteins names associated with all HGNC symbols mentioned in this chapter have been listed in Table 5.9 at the end of this chapter.							

Table 5.4: **Top 10 upregulated and downregulated DE proteins for A0 v S0 comparisons ranked by pvalue and Log_2FC value from Figure 5.7.**

Volcano plots (Figures 5.6 and 5.7) were used to visually represent which DE proteins were upregulated or downregulated for each comparison.

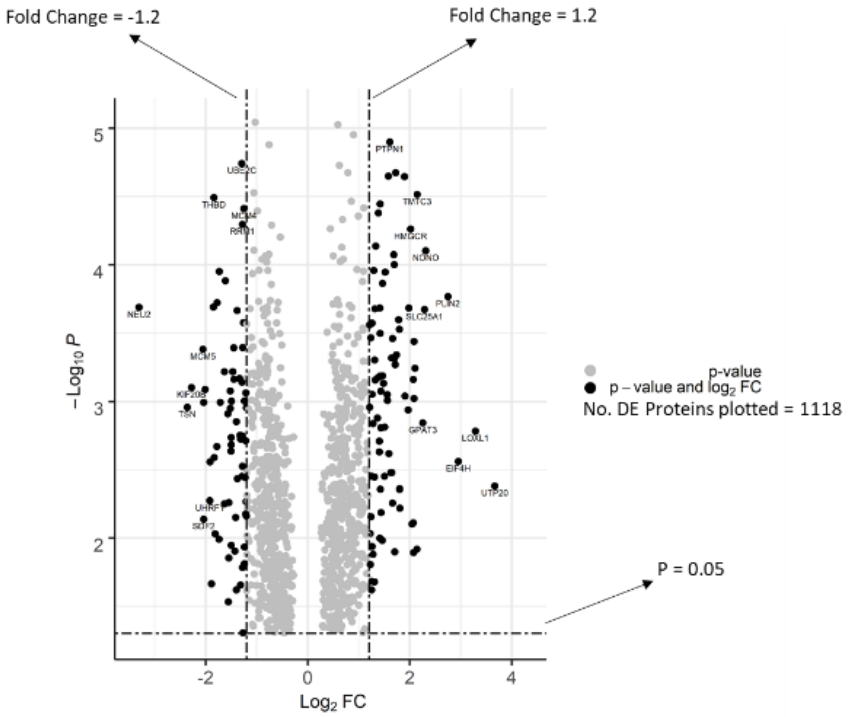


Figure 5.6: Volcano plot showing the distribution of DE proteins for the comparison of attached and suspension CHO K1 cultures grown in media containing 5% serum (A5vS5). Black dots represent DE proteins with a FC > ±1.2 and a p <0.05. Grey dots represent DE proteins with an FC < ± 1.2 and a p <0.05.

Each individual point on a volcano plot represents the fold change between two experimental groups plotted against the level of statistical difference. To obtain the best trade-off between having too much data for analysis and the risk of excluding data the FC cut off for the volcano plot analysis was set at 1.2 as an FC of 1.2 was used during data processing with Progenesis QI for Proteomics software. The use of volcano plots aided interpretation of whether the comparison resulted in DE proteins which were predominantly upregulated (positive FC) or downregulated (negative FC). Most of the DE proteins were upregulated when CHO K1 attached and suspension cultures grown in the absence of serum were compared (one hundred and eighty-eight downregulated and four hundred and eighteen upregulated), whereas the distribution was nearly equal when CHO K1 attached and suspension cultures grown in media containing 5% FBS were

compared (two hundred and fifty-two downregulated and two hundred and eight-seven upregulated).

Figure 5.6 shows the fold change distribution of one thousand one hundred and eighteen DE proteins identified in the A5 v S5 experimental group. The most significantly changed protein was the small subunit processome component 20 homolog (UTP20, Figure 5.6). It had a fold change of 12.8 and its expression level was highest in 5% suspension culture. The expression of UTP20 was upregulated in suspension cells compared to attached cells as it's located on the right hand side of the volcano plot. A role has been described for this protein in 18S pre-rRNA processing and it is known to associate with U3 snoRNA (Wang et al, 2007). According to UniProtKB, UTP20 also has a role in the negative regulation of cell population proliferation. Seven of the top ten changed proteins had their highest expression levels in suspension culture. The 3 proteins which had the greatest change in expression, in attached culture, were *sialidase-2* (NEU2, FC = 7.7, Figure 5.5), *the DNA binding protein, translin* (TSN, FC = 4.3, Figure 5.6) and *kinesin-like protein KIF20B* (KIF20B, FC = 4.3, Figure 5.6). These all appeared on the left hand side of the volcano plot and therefore their expression was downregulated in attached cells compared to suspension cells. According to UniProtKB, Sialidase-2 is known to play a role in oligosaccharide catabolism, namely the removal of the terminal sialic acid from glycolipids, glycoproteins and oligosaccharides. Sawada et al. (2002) reported that when the rat *Neu2* gene was transfected into and expressed in the highly metastatic mouse colon 26 adenocarcinoma cell line, a reduction in lung metastasis, invasion and mobility was observed. These observations correlated with a decrease in the levels of sialyl Le^x and the ganglioside GM3. Additionally, the authors showed that by incubating the cell line with antibodies against sialyl Le^x and GM3 cell adhesion was inhibited compared to controls. This observation provided direct evidence that the desialylation of sialyl Le^x and GM3 by sialidase is involved in the suppression of metastasis.

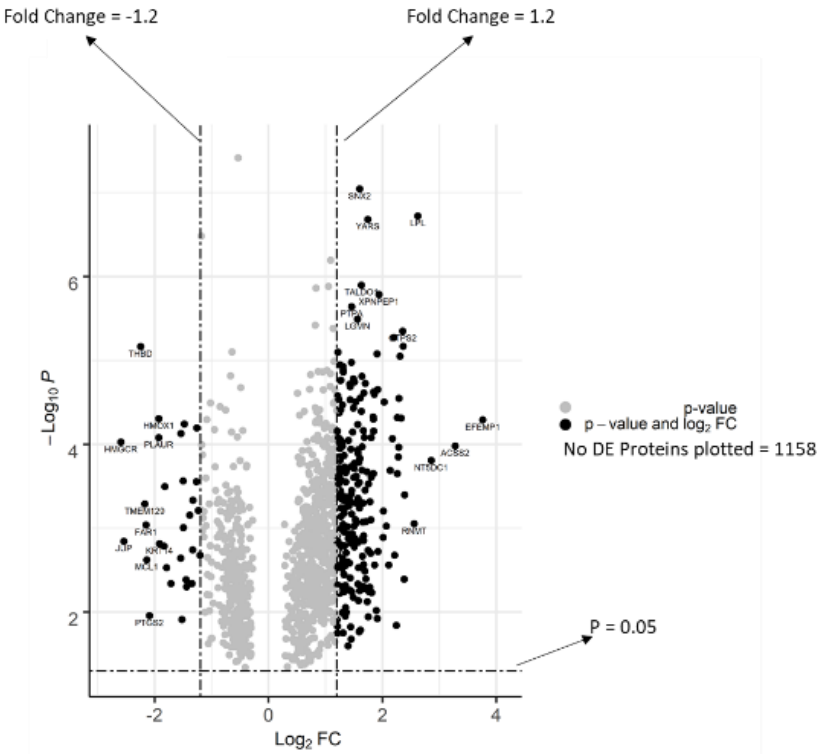


Figure 5.7: **Volcano plot showing the distribution of DE proteins for the comparison of attached and suspension CHO K1 cultures grown in media without serum (A0vS0). Black dots represent DE proteins with a FC > ±1.2 and a p <0.05. Grey dots represent DE proteins with an FC < ± 1.2 and a p <0.05.**

Figure 5.7 shows the fold change distribution of one thousand one hundred and fifty-eight DE proteins identified in the A0 v S0 experimental group. The most significantly changed protein was *EGF-containing fibulin-like extracellular matrix protein 1* (EFEMP1, Figure 5.7). It had a fold change of 13.3 and its expression level was highest in suspension culture. The expression of EFEMP1 was upregulated in suspension cells compared to attached cells as it's located on the right hand side of the volcano plot. Hu et al. (2019) showed that reduced expression of EFEMP1 was correlated with a reduction in metastasis of hepatocellular carcinoma. They proposed that the interaction between increased expression of EFEMP1 and the subsequent increased expression of a second protein, *Semaphorin-3B precursor* (SEMA3B), which was also present in this dataset was the mechanism by which metastasis was reduced. Interestingly *SEMA3B*, was present in both the 5% FBS and 0% FBS comparisons. The fold change in 0% suspension culture was 5.0, while the fold change in 5% suspension was 1.7.

Eight of the top ten changed proteins had their highest expression levels in suspension culture. The 2 proteins which were most abundant in attached culture were *3-hydroxy-3-methylglutaryl-coenzyme A reductase* (HMGCR, Figure 5.7) and j*unction plakoglobin* (JUP, FC = 5.8 Figure 5.7). Both 3-hydroxy-3-methylglutaryl-coenzyme A reductase and junction plakoglobin were downregulated in attached cells compared to suspension cells since they are both located on the left-hand side of the volcano plot (Figure 5.7). HMGCR was detected in both the 5% FBS and 0% FBS comparisons. The fold change in 0% attached culture was 6.0, while the fold change in 5% suspension was 4.0. HMGCR catalyses the production of mevalonate from *3*-hydroxy-3-methylglutaryl-CoA, a critical step in the synthesis of cholesterol. Inhibition of this reaction prevents the synthesis of other important isoprenoid intermediates of the cholesterol biosynthetic pathway which function as binding sites for a number of proteins including the small guanosine triphosphate (GTP) binding proteins Ras. Ras proteins have been reported as being involved in cell proliferation, apoptosis, and most importantly, cell adhesion (Sandri et al., 2012). Junction plakoglobin has been shown to play a role in cell-cell adhesion and the formation of adherens junctions via the stimulation of VE-cadherin function in endothelial cells.

5.3.3. PSEA analysis of DE proteins detected in the A5vS5 and A0vS0 experimental groups

PSEA analysis was used to identify significantly enriched KEGG pathways, the DE proteins implicated in those pathways and the upregulated or downregulated status of those DE proteins for both the A5vS5 and A0vS0 datasets. PSEA analysis of all one thousand one hundred and eighteen DE proteins in the A5vS5 dataset (Appendix 3) revealed one hundred and fourteen significantly enriched ($p \leq 0.05$) KEGG pathways involving two hundred and fifty-two upregulated and two hundred and eighty-seven downregulated DE proteins (Appendix 4). Similarly, PSEA analysis of all one thousand one hundred and fifty-eight DE proteins in the A0vS0 dataset (Appendix 5) revealed one hundred and forty-eight significantly enriched pathways (Appendix 6) involving four hundred and eighteen upregulated and one hundred and eighty-eight downregulated DE proteins.

The top ten enriched KEGG pathways based on fold change are presented in Table 5.5 and 5.6 respectively.

ID	KEGG Pathway	Fold Enrichment	P value	Downregulated	Upregulated
hsa04066	HIF-1 signalling pathway	2.99	$7.61E10^{-05}$	TFRC, HMOX1, SLC2A1, PDHA1, PDHB	MAP2K1, MAPK1, EIF4E, PLCG1, CAMK2D, HK1, PFKL, GAPDH, ALDOA, ENO1, PGK1, LDHA
hsa04723	Retrograde endocannabinoid signalling	2.59	$8.29E10^{-16}$	PTGS2, GNAI3, GNAI2, NDUFV1, NDUFV2, NDUFA5, NDUFA9, NDUFA10, NDUFA13, NDUFB5, NDUFB7, NDUFB8, NDUFB9, NDUFB10, NDUFB11, NDUFS1, NDUFS2, NDUFS5, NDUFS8	MAPK1
hsa04922	Glucagon signalling pathway	2.58	0.0008	GNAS, SLC2A1, PDHA1, PDHB	PPP3CA, PPP3R1, CAMK2D, PCK2, PYGB, ACACA, PFKL, PGAM1, PKM, LDHA
hsa04370	VEGF signalling pathway	2.55	0.005	PTGS2, SRC	PLCG1, MAP2K1, MAPK1, PPP3CA, PPP3R1, HSPB1
hsa04915	Estrogen signalling pathway	2.33	$1.14E10^{-6}$	KRT10, KRT14, KRT17, GNAS, SRC, MMP9, GNAI3, GNAI2	HSP90AA1, HSP90AB1, FKBP4, HSPA8, HSPA2, CTSD, GRB2, MAP2K1, MAPK1
hsa04910	Insulin signalling pathway	1.92	0.001	None	PYGB, PRKAR1A, PRKAR2B, FLOT2, FLOT1, CRK, ACACA, FASN, HK1,

ID	KEGG Pathway	Fold Enrichment	P value	Downregulated	Upregulated
hsa04910	Insulin signalling pathway	1.92	0.001	None	PCK2, EIF4E, GRB2, MAP2K1, MAPK1
hsa04921	Oxytocin signalling pathway	1.88	0.0003	PTGS2, CACNA2D1, ACTB, GNAS, GNAI3, GNAI2, SRC	MAP2K1, MAPK1, EEF2, PPP3CA, PPP3R1, CAMK1, CAMK2D, PPP1R12A
hsa04022	cGMP-PKG signalling pathway	1.87	0.0002	GNA13, ATP1A1, ATP1B3, ATP2B1, ATP2A2, GNAI3, GNAI2, VDAC1, VDAC2, VDAC3, SLC25A5	PPP3CA, PPP3R1, PPP1R12A, MAP2K1, MAPK1
hsa04919	Thyroid hormone signalling pathway	1.86	0.016	ITGAV, ATP2A2, ATP1A1, ATP1B3, SRC, SLC2A1, ACTB	MAP2K1, MAPK1, HDAC2, PLCG1, PFKL
hsa04012	ErbB signalling pathway	1.81	0.003	SRC	PLCG1, CAMK2D, CRK, PAK2, GRB2, MAP2K1, MAPK1
Protein names have not been used for space reasons. Instead, the proteins names associated with all HGNC symbols mentioned in this chapter have been listed in Table 5.9 at the end of this chapter.					

Table 5.5: **Top 10 enriched KEGG pathways and the DE proteins detected in those paths as identified by PSEA analysis of DE proteins detected in the A5vS5 experimental group.**

ID	KEGG Pathway	Fold Enrichment	P value	Downregulated	Upregulated
hsa04066	HIF-1 signalling pathway	3.71	3.23E-05	TFRC, HMOX1, SLC2A1, PDHA1	MAPK1, EIF4E, PLCG1, CAMK2D, HK1, PFKP, PFKL, GAPDH, ALDOA, ENO1, PGK1, LDHA
hsa04922	Glucagon signalling pathway	3.69	0.04	GNAQ, SLC2A1, PDHA1	PPP3CA, PPP3R1, CAMK2D, PCK2, PYGM, PYGB, ACACA, PFKP, PFKL, PGAM1, PKM, LDHA
hsa04910	Insulin signalling pathway	2.56	0.0006	None	PYGM, PYGB, PRKAR1A, PRKAR2B, FLOT2, FLOT1, CRK, ACACA, FASN, HK1, PCK2, EIF4E, GRB2, MAPK1
hsa04723	Retrograde endocannabinoid signalling	2.55	1.66E-10	GNAQ, PTGS2, GNAI2, NDUFV2, NDUFA5, NDUFA10, NDUFA13, NDUFB7, NDUFB8, NDUFB9, NDUFS1, NDUFS2, NDUFS5, NDUFS8	MAPK1
hsa04370	VEGF signalling pathway	2.52	0.03	PTGS2	PLCG1, MAPK1, PPP3CA, PPP3R1, HSPB1
hsa04915	Estrogen signalling pathway	2.17	0.02	KRT10, KRT14, KRT17, MMP9, GNAI2, GNAQ	HSP90AB1, FKBP4, HSPA2, CTSD, GRB2, MAPK1
hsa04152	AMPK signalling pathway	2.10	0.02	HMGCR	PFKP, PFKL, PCK2, EEF2, FASN, ACACA, PPP2R1A, PPP2R2A, CD36
hsa04921	Oxytocin signalling pathway	1.99	0.04	GNAQ, PTGS2, CACNA2D1, ACTB, GNAI2	MAPK1, EEF2, PPP3CA, PPP3R1, CAMK1, CAMK2D, PPP1R12A

ID	KEGG Pathway	Fold Enrichment	P value	Downregulated	Upregulated
hsa04071	Sphingolipid signalling pathway	1.91	0.04	SPTLC1, GNAI2, GNAQ, GNA13	ASAH1, CTSD, PPP2R1A, PPP2R2A, MAPK1
hsa04012	ErbB signalling pathway	1.79	0.04	None	PLCG1, CAMK2D, CRK, PAK2, GRB2, MAPK1
Protein names have not been used for space reasons. Instead, the proteins names associated with all HGNC symbols mentioned in this chapter have been listed in Table 5.9 at the end of this chapter.					

Table 5.6: **Top 10 enriched KEGG pathways and the DE proteins detected in those paths as identified by PSEA analysis of DE proteins detected in the A0vS0 experimental group.**

5.3.4. Gene Ontology analysis of DE proteins detected in the A5vS5 and A0vS0 experimental groups

Gene ontology analysis of all differentially expressed proteins, which were present in the significantly enriched ($p \leq 0.05$) KEGG pathways identified by PSEA analysis, in the A5 v S5 dataset revealed an association with 281 GO terms (Appendix 7) which could be grouped into thirty-two separate biological processes (Figure 5.8). The discrepancy in numbers could be explained by the fact that a single protein was associated with multiple GO terms and played a role in multiple biological processes. For example, MCM5 was associated with four distinct GO numbers describing seven different biological processes. The top five categories of biological processes, accounting for 57% of all enriched biological processes, which resulted from a proteomic differential expression analysis comparing CHO K1 cells grown as attached or suspension cells in the presence of serum, are listed in Table 5.7.

Leaving aside the catch-all terms "biological process", "metabolic process" and GO:0008283 – cell proliferation, the remaining 29 GO terms were further categorised into five broad themes. These were:

- Production of energy – GO:0006091, GO:0007005, GO:0005975
- Metabolism – GO:15031, GO:0006629, GO:0006464, GO:0009058, GO:0006810, GO:0006811
- Cell Cycle/Growth – GO:0007049, GO0008283, GO:0008219, GO:0007275, GO:0009790, GO:0009653, GO:0007010, GO:0007028, GO:0016049, GO:0019725, GO:0030154
- Generation/replication of genetic material - GO:000635, GO:0006139, GO:0006259, GO:0006412, GO:0016032
- Cell Recognition/Communication – GO:0007165, GO:0007154, GO:0007267, GO:0008037

G0:0008283 – cell proliferation was excluded as its definition according to the gene ontology website (https://www.ebi.ac.uk/QuickGO/term/GO:0008283) is "The multiplication or reproduction of cells, resulting in the expansion of a cell population.

Note that this term is intended to be used for the proliferation of cells within a multicellular organism, not for the expansion of a population of single-celled organisms".

Themes such as the production of energy, metabolism, communication, cell cycle and the generation/replication of genetic material are common to all living cells. However, GO terms such as morphogenesis, cytoskeleton organisation, cell recognition, cell differentiation, cell-cell signalling and cell communication describe processes central to the adaption from attached growth to growth in suspension. The GO term "cell recognition" covers processes whereby a cell responds to changes in its environment, e.g., adaptation to suspension growth. The epithelial-mesenchymal transition is a well described cell differentiation process under which polarised epithelial cells detach from a surface and assume a mesenchymal phenotype and which plays a key role in the metastasis of cancer cells. Cell-cell signalling and cell communication involve intercellular contact via structures such as gap junctions. It's reasonable to assume these types of cellular attachments will be disrupted during the adaption to suspension growth. GO terms associated with "morphogenesis" and "cytoskeletal organisation" cover the structural changes that occur as cells reorganize their physiology away from cell-cell and cell-surface interactions towards suspension growth. The GO term GO:0009653 which is related to morphogenesis is the only term in this list which is not also present in the A0 v S0 comparison, perhaps indicating that morphological changes are among the first to occur when cells adapt from attached to suspension growth. Finally, it would be naïve to think that all cells seamlessly adapt from attached growth to suspension growth. Hence the appearance of GO terms such as the response to stress and cell death seem appropriate. The adaptation process is a long and slow process characterised by periods of low cell growth and viability as cells that failed to adapt to suspension growth die.

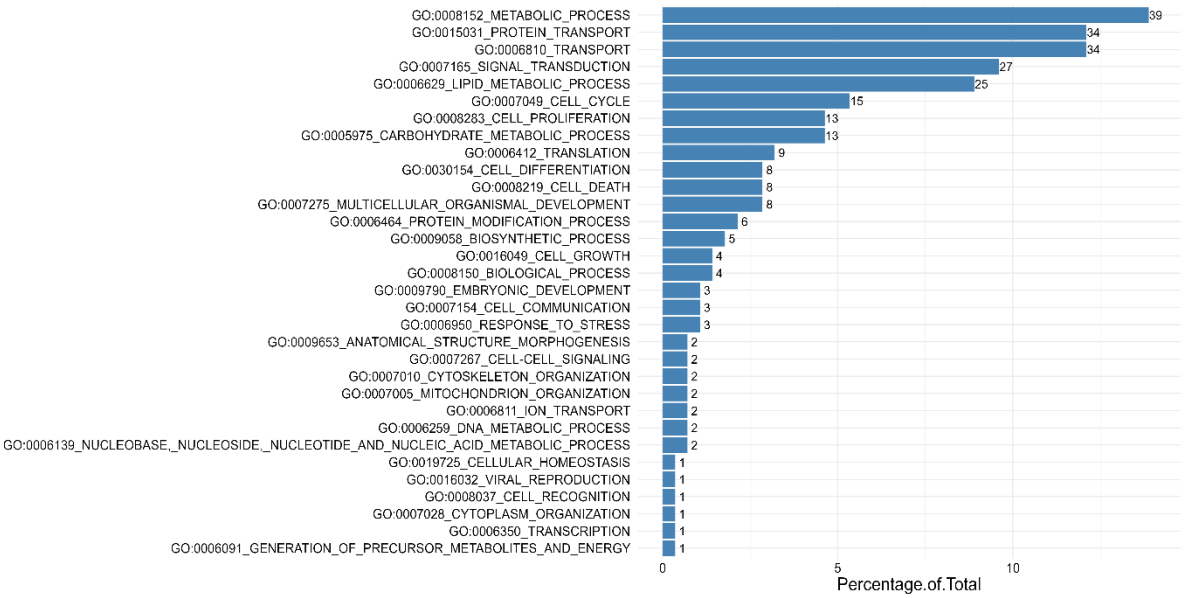


Figure 5.8: **Categorisation of the enriched GO biological process terms related to DE proteins identified in the A5 v S5 experimental group.**

GO Category	No. of GO Terms	% of the Total	Top 5 Cumulative percentage
GO:0008152 METABOLIC PROCESS	39	13.88	56.59
GO:0015031 PROTEIN TRANSPORT	34	12.10	
GO:0006810 TRANSPORT	34	12.10	
GO:0007165 SIGNAL TRANSDUCTION	27	9.61	
GO:0006629 LIPID METABOLIC PROCESS	25	8.90	

Table 5.7: **Top five categories of biological process the enriched biological process GO terms from Figure 5.8 (A5 v S5 experimental group) fall into.**

Gene ontology analysis of all differentially expressed proteins, which were present in the significantly enriched ($p \leq 0.05$) KEGG pathways identified by PSEA analysis, in the A0 v S0 dataset revealed an association with 345 GO terms (Appendix 8) which could be grouped into thirty-three separate biological processes (Figure 5.9). The top five categories of biological processes, accounting for 53% of all enriched biological processes, which resulted from a proteomic differential expression analysis comparing

CHO K1 cells grown as attached or suspension cells in the presence of serum, are listed in Table 5.8.

Comparison of the thirty-three GO terms from the A0 v S0 comparison (Figure 5.9) with those obtained from the A5 v S5 comparison (Figure 5.8) revealed two additional terms. These were GO:0019538, which covers metabolic processes involving proteins, and GO:0009719, which covers the response to endogenous stimuli. It's known that CHO K1 cells grow slower in suspension that they do as attached culture. It's likely that slowdown is caused by changes in the cells metabolism as it adapts to suspension growth. Considering the changes (e.g., loss of adhesion) that cells undergo when adapting to suspension growth it's not surprising that proteins involved in the cellular response to endogenous stimuli are differentially expressed. The top five categories of biological processes, accounting for 53% of all enriched biological processes, which resulted from a proteomic differential expression analysis comparing CHO K1 cells grown as attached or suspension cells in the absence of serum, are listed in Table 5.8.

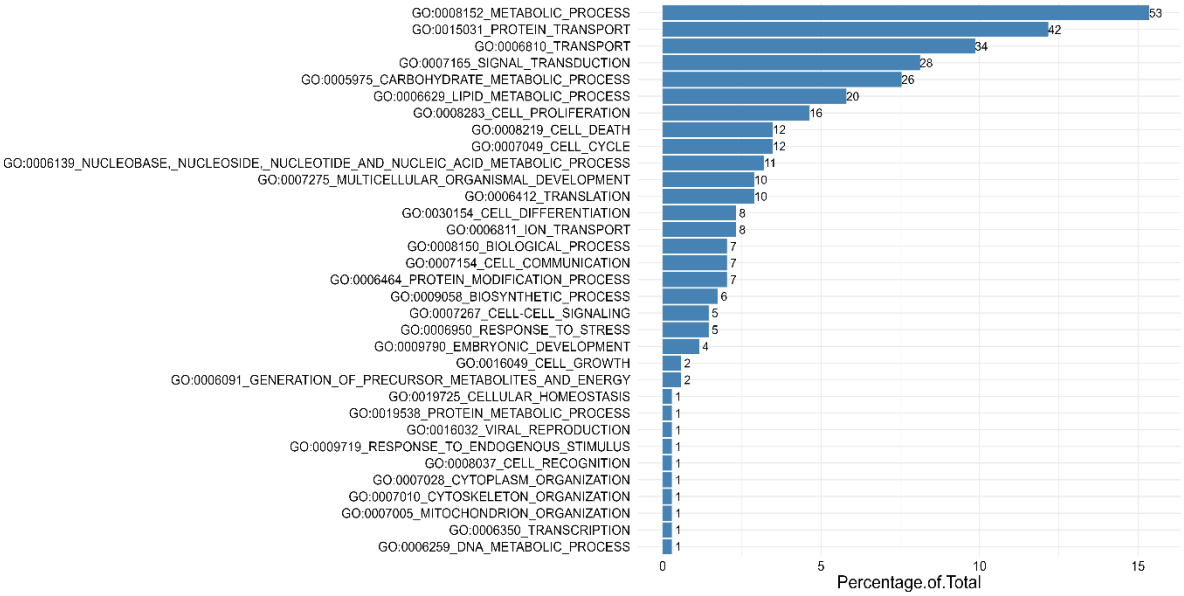


Figure 5.9: **Categorisation of the enriched GO biological process terms related to DE proteins identified in the A0 v S0 experimental group.**

GO Category	No. of GO Terms	% of the Total	Top 5 Cumulative percentage
GO:0008152 METABOLIC PROCESS	53	15.36	53.04
GO:0015031 PROTEIN TRANSPORT	42	12.17	
GO:0006810 TRANSPORT	34	9.86	
GO:0007165 SIGNAL TRANSDUCTION	28	8.12	
GO:0005975 CARBOHYDRATE METABOLIC PROCESS	26	7.54	

Table 5.8: **Top five categories of biological process the enriched biological process GO terms from Figure 5.9 (A0 v S0 experimental group) fall into.**

5.4. Discussion

This chapter discusses the data resulting from a comparison of CHO K1 cells grown attached to the surface of a flask and those grown in suspension. Lists of differentially expressed proteins were prepared using Progenesis QI for Proteomics software. Considering that the major difference between the two cell lines was their way of growing it's reasonable to expect that proteins related to attachment/adhesion would be present. Attachments between cells can be divided between cell-cell contacts, e.g., tight and adherens junctions, attachment of the cell to a surface by anchoring junctions e.g., desmosomes and hemidesmosomes and gap junctions. Gap junctions tend to be involved in intracellular communication rather than adhesion and thus aren't of interest in this study. Broadly speaking the adhesion molecules which mediate these attachments are categorised into 4 groups. These are cadherins, integrins, CD44 and the immunoglobulin superfamily. Typically, tight junctions involve members of the immunoglobulin super family known as junctional adhesion molecules (JAM) and multiple plaque proteins. On the other hand, adherens junctions and desmosomes involve cadherins such as E-cadherin, desmogleins and desmocollins while anchoring junctions such as hemidesmosomes are characterised by integrins. Therefore, it was expected that changes in the expression levels of these types of proteins would be observed or that biological processes involving these types of attachments would be

enriched, in addition to processes involving cell proliferation, cell differentiation. However, when attached and suspension cells grown in the presence of FBS (A5vS5) or absence of FBS (A0vS0) were compared cell differentiation and cell proliferation/adhesion didn't dominate the list of enriched GO processes.

5.4.1. Reproducibility of the number of proteins identified in each replicate and the impact on the identification of enriched KEGG pathways

In excess of three thousand proteins were detected in each of the replicates in both the A5vS5 and A0vS0 experimental groups. However, not all these proteins were detected in all replicates and some of the potential reasons for this are mentioned in Section 5.3.1. Table 5.2 lists the proteins present in enriched KEGG pathways which were not common to all replicates. The pathway with the greatest number of DE proteins which weren't detected in all replicates was the Estrogen signalling pathway with five proteins (three down regulated and two up regulated). However, given the total number of proteins detected in this pathway (eight down regulated and nine up regulated) it's unlikely that the fact that five proteins were detected as being differentially expressed while not being detected in all replicates influenced the identification of the Estrogen signalling pathway as being enriched. This is based on the observation that the VEGF signalling pathway was detected as enriched despite only five DE proteins being associated with this pathway. This reasoning can also be applied to the other pathways discussed in Sections 5.4.2 and 5.4.3 and as such it's considered unlikely that the variability in the total number of proteins detected in each replicate (Table 5.1) impacted on the KEGG paths identified as differentially expressed in Tables 5.5 and 5.6.

Table 5.3 and Table 5.4 list the top 10 ten upregulated and downregulated DE proteins based on their p value and fold change for the A5vS5 and A0vS0 comparisons respectively. Except for GPAT3 (*Glycerol-3-phosphate acyltransferase 3)* and CRNKL1 (*Crooked neck pre-mRNA splicing factor 1*) the proteins listed in Table 5.3 were detected in all replicates of one experimental group and at least 2 replicates of the second experimental group. Additionally, neither *Glycerol-3-phosphate acyltransferase 3* nor

Crooked neck pre-mRNA splicing factor 1 were detected in any of the enriched KEGG signalling pathways listed in either Table 5.5 or Table 5.6.

With respect to Table 5.4 FAR1 (*fatty acyl-CoA reductase 1*) and TMEM129 (*E3 ubiquitin-protein ligase TM129*) were both detected in three replicates of one experimental group and only one replicate of the second experimental group and therefore are considered to be differentially expressed. MCL1 (*Induced myeloid leukaemia cell differentiation protein MCL-1)* can't be considered as a true differentially expressed protein as it was only detected in two replicates in total. JUP (*Junction plakoglobin*) and PTGS2 (*Prostaglandin-endoperoxide synthase 2*) were only detected in a single replicate and therefore can't be considered as true differentially expressed proteins. PTGS2 was detected in some of the enriched KEGG signalling pathways listed in Table 5.5 and Table 5.6. However, given that multiple DE proteins in addition to PTGS2 were detected in these pathways it's unlikely that removing PTGS2 from the analysis would impact on the enriched status of the KEGG pathways listed in Table 5.5 and Table 5.6. Performing multiple injections from the same MS sample vial (technical replicates) could help to establish whether these proteins were truly differentially expressed. However, technical replicates were not performed due to the number of samples for analysis and the injection time required for each sample.

5.4.2. Enriched signalling pathways detected in A5vS5 experimental group by PSEA

Examination of the data presented in Figure 5.8 shows that cell proliferation as a category ranked 7th and accounted for 4.63% of the enriched biological processes. Cell differentiation as a category ranked 13th and accounted for 2.85% of the enriched biological processes and while adhesion as a GO term was enriched and several adhesion pathways were enriched, adhesion as a category was not. The answer to the apparent contradiction may derive from the fact that signal transduction was found to be the fourth most significantly enriched category accounting for 9.61% of the enriched GO terms. Signal transduction is known to be a key step in the control of cellular functioning. It describes the mechanism by which signals are transmitted throughout the cell,

resulting in cellular processes being turned on and off. PSEA analysis identified twenty-two separate signalling pathways which were found to be enriched in this study.

Three of the most significant from a fold change viewpoint were:

1. The HIF-1 signalling pathway;
2. The glucagon signalling pathway; and
3. The VEGF signalling pathway.

In all cases the upregulation of a protein has been defined as when that proteins expression in suspension adapted cells is higher that it's expression in attached cells.

HIF stands for hypoxia-inducible factor (reviewed by Brahimi-Horn and Pouysségur (2009) and Lee et al., (2019). Hypoxia has long been associated with metastasis of cancer cells (reviewed by Hill et al., 2009). An early step in metastasis, is a loss of adhesion between individual cancer cells and the primary tumour. It therefore makes sense that despite the fact that the CHO cells in this study were not exposed to hypoxic conditions HIF signalling would be enriched due to its association with loss of cellular adhesion. It is known that cancer cells which metastasise often form clusters or aggregates (Aceto et al., 2014) and this propensity to form cell aggregates is thought to be one of the mechanisms by which metastasising cancer cells survive after detachment from the ECM. Jiang et al. (2016) linked this ability to survive detachment with increased NADPH production mediated by the enzyme isocitrate dehydrogenase 1 (IDH2) and a switch from oxidative phosphorylation (OXPHOS) towards glycolysis. This observation was supported by Schafer et al (2009) who demonstrated that while a reduced flux through OXPHOS lead to increased ROS production and cell death, cell survival was improved by promoting glycolysis. Labuschagne et al. (2019) demonstrated that this metabolic switch from OXPHOS toward glycolysis (as characterised by increased reductive carboxylation) was supported by the induction of hypoxia in the aggregated cells. They ruled out that the induction of hypoxia was due to limited oxygen diffusion through the aggregates and demonstrated that it was due to the formation of cell aggregates. The aggregated cells also underwent mitophagy, induced by increased expression of Hif1a, BNIP3 and NIX(BNIP3L). BNIP3 and NIX are both the products of genes known to be targeted by

Hif1a. The final consequence of the formation of cell aggregates was a reduction in ROS due to increased mitophagy and improved cell survival compared to detached cells which were prevented from aggregating by the inclusion of EDTA or the drug ouabain in the culture media. Interestingly the glycolytic enzymes, *hexokinase, phosphofructokinase, GAPDH, phosphoglycerate kinase, enolase* and *lactate dehydrogenase* were all shown to be upregulated when comparing suspension adapted CHO K1 cells which were grown in attached and suspension culture.

The glucagon signalling pathway plays a role in maintaining blood glucose homeostasis in the body. It works by promoting the catabolism of glucose. There is evidence that glucagon like peptide receptor agonists can inhibit adhesion of monocytes to endothelial cells (Arakawa et al., 2010). However, what's more interesting in the context of adhesion is the finding that glucose itself can control the expression of adhesion molecules. Baumgartner-Parzer et al. (1995) showed that increased glucose concentration correlated with increased expression of ICAM-1 in HUVECs and that this effect was time dependent, i.e., the longer cells were incubated in the presence of glucose the greater the increase in ICAM-1 expression was. Srinivasan et al. (2003) reported a similar observation in human aortic endothelial cells (HAECs). They showed that growing HAECs in the presence of glucose correlated with an increase in IL-8 expression and a consequent activation of the β1 integrin on the surface of HAECs. They also correlated increases in glucose concentration with increased adhesion of monocytes to HAECs. Incubation of HAECs with an anti-IL-8 antibody inhibited this effect and reduced the ability of monocytes to adhere to HAECs. Obviously in the context of this study the observation that the *glucose transporter 1 protein*, encoded by SLC2A1, was down regulated is significant as it's the protein which regulates entry of glucose into the cell and therefore controls this entire pathway.

The vascular endothelial growth factor (VEGF) signalling pathway is a key pathway in the formation and growth of blood vessels. Additionally, a role for VEGF has been demonstrated in the promotion of cell migration, proliferation, and survival. One of the proteins present in the VEGF signalling pathway which was detected in this study was the small heat shock protein B1 (HSPB1). In 1997, Lemieux et al. (1997) showed that

HSPB1 overexpression enhanced adhesion in vitro. Subsequently several reports appeared in the literature describing a role for HSPB1 in cell adhesion in vivo. Fanelli et al. (2008) demonstrated that overexpression of HSPB1 resulted in β-catenin relocating from the cell membrane to the cytoplasm. Following relocation to the cytoplasm, β-catenin interacted with HSPB1 and controlled the expression of cadherin-catenin cell adhesion proteins.

5.4.3. Enriched signalling pathways detected in A0vS0 experimental group by PSEA

Comparison of DE proteins from attached and suspension cells grown in the absence of serum showed a similar picture to those grown in the presence of serum. Cell differentiation and cellular proliferation/adhesion as categories didn't dominate the list of enriched biological processes when cells grown as attached or in suspension in the absence of serum were compared. Cell proliferation as a category ranked 7th and accounted for 5% of the enriched biological processes. Cell differentiation as a category ranked 13th and accounted for only 2% of the enriched biological processes. Moreover, while adhesion as a GO term was enriched and several adhesion pathways were enriched, adhesion as a category was not.

The answer to this contradiction may once again be found in the fact that signal transduction was found to be the 4th most statistically significantly enriched category, accounting for 8% of the enriched GO terms. Enriched KEGG pathway analysis identified twenty-six separate signalling pathways which were found to be enriched in this study. The three most significant from a fold change viewpoint which were not enriched in the 5% FBS data were:

1. Insulin signalling pathway;
2. Estrogen signalling pathway; and
3. AMPK signalling pathway.

In all cases the upregulation of a protein has been defined as when that proteins expression in suspension adapted cells is higher that it's expression in attached cells.

It has been shown that insulin plays a vital role in the adhesion of human embryonic stem cells (Godoy-Parejo et al., 2109). Additionally, Wolf et al. (2013) demonstrated that insulin signalling played a role in switching the function of plakophilin1 from mediation of cellular adhesion to cellular proliferation. It's therefore not surprising that the insulin signalling pathway was significantly enriched in this dataset. Among the proteins involved in this pathway, the *growth factor receptor bound protein 2* (GRB2) was found to be upregulated. Cheng et al. (2014) demonstrated that GRB2 promoted the induction of focal adhesion kinase (FAK) phosphorylation. Given the well understood role the phosphorylation of FAK plays in the disruption of cellular attachment, it's reasonable to suggest that the upregulation of GRB2 expression played a role in the adaptation of CHO K1 cells from attached growth to suspension growth. Another protein present in this dataset from the insulin signalling pathway was *eukaryotic translation initiation factor 4E*, which has been shown to be implicated in the adhesion and pseudohyphal growth of *S. cerevisiae* cells. Ross et al. (2012) established a correlation between eukaryotic translation initiation factor 4E expression levels and adhesion. A reduction in eukaryotic translation initiation factor 4E expression positively correlated with a reduction in adhesion. In this study EIF4E was found to have lower expression in attached cells compared to suspension cells.

The AMPK signalling pathway has been described as "one of the central regulators of cellular and organismal metabolism in eukaryotes" (reviewed by, Mihayolva and Shaw, 2011). As well as regulating cell growth and metabolism (fatty acid synthesis) roles have been described for it in regulating adhesion and cell polarity. The protein phosphatase 2A (PP2A) is a multi subunit protein, two of which were detected as being upregulated (highest expression in suspension cells) in this study. In their review of cell adhesion and the role PP2A plays in it Sontag and Sontag (2006) describe the mechanism by which PP2A disrupts focal adhesions and promotes cell migration.

The AMPK signalling pathway also incorporates many enzymes involved in lipid metabolism. Acetyl-CoA carboxylase alpha (ACACA), fatty acid synthase and 3-hydroxy-3-methylglutaryl CoA reductase (HMGCR) are enzymes involved in lipid metabolism which were shown to be implicated in the AMPK signalling pathway which were found

to be differentially expressed in this study. ACACA catalyses the carboxylation of acetyl-CoA into Malonyl CoA. This is the rate limiting step in the biosynthesis of all fatty acids. The enzyme responsible for the first step in the synthesis of cholesterol, the conversion of HMG-CoA into mevalonate, 3-hydroxy-3-methylglutaryl CoA reductase (HMGCR) was also detected in this study. ACACA was found to have its highest expression in suspension cells and was therefore upregulated. Conversely HMGCR was found to be downregulated in this study (highest expression in attached). Additionally3-hydroxy-3-methylglutaryl CoA reductase. Additionally, the fatty acid transport protein, CD36 was found to be upregulated in this study (highest expression in suspension cells). Xu et al (2013) described a role for CD36, in HEK cells, in increasing the rate at which fatty acids were incorporated into triglycerides independently of the rate at which fatty acids were transported across the plasma membrane.

One potential advantage of this increased fatty acid biosynthesis is that it allows the cell to build up reserves of fatty acids which can be called upon as a source of ATP and NADPH should the cell get metabolically stressed (Petan et al., 2018). Given the critical role NADPH plays in the detoxification of ROS this enrichment of lipid biosynthesis by the AMPK signalling pathway maybe a method by which the cell protects itself and enhances its survival during adaptation to suspension growth. A second potential reason for enhanced lipid accumulation is the impact this may have on membrane fluidity. Hu et al. (2011) reviewed strategies for minimising the possibility for hydrodynamic damage to cells in bioreactors. While they recognise that the perception that high impeller rpm rates are damaging to cells has its origins in the use of microcarriers during suspension culture rather than current practice it's probably an erroneous perception given the success the biologics industry has had in developing suspension cultures in 20,000L reactors. Accordingly, it's interesting to speculate that increased fatty acid biosynthesis by suspension adapted cells lead to increased membrane fluidity and that this bestowed a protective effect on cells with regards to the shear forces encountered in bioreactors compared to attached cells cultivated in suspension on microcarriers.

The fact that proteins linked to metastasis of cancer have been identified as differentially expressed in this analysis is not surprising from an attachment point of view. Metastasis

is the process whereby cancer cells migrate from the primary tumour and seed secondary tumours at distant locations. Consideration of the events involved in this process leads to the conclusion that the detachment and reattachment of cells are key. Cells must initially detach from the primary tumour, travel through the body, reattach and start to grow at the new site. Multiple signalling pathways have been implicated in metastasis including wnt/β-catenin (Zhang et al, 2018), hypoxic (Rankin et al, 2016), hippo (Janse van Rensburg and Yang, 2017), VEGF (Liu et al, 2011), notch (Hu et al, 2012) and sonic hedgehog (Yoo et al, 2011). Xie and Huang (2003) reviewed the role that cellular stress pathways have in regulating metastasis of cancer cells. Two of the pathways they reviewed as being capable of promoting metastasis were hypoxia, as regulated by HIF-1, and metabolic pathways which control the cellular response to extremes of glucose. Proteins involved in the cellular response to hypoxia and fluctuations in glucose concentration were found to be differentially expressed in this study.

5.5. Relevance of this study to CHO cell culture:

This study compared the proteome of CHO K1 cells grown as both attached and suspension cells in the presence and absence of serum. It has provided insights into the key biological pathways that potentially underscored the development of suspension adapted CHO K1 cell lines. More specifically, the detection of multiple differentially expressed proteins, which are known to be implicated in various cell signalling pathways, and their role in regulating cellular attachment was highlighted.

Taking this data, it's clear that the approach described has identified multiple proteins endogenous to CHO cell lines which are potential targets for engineering with the aim of developing a CHO cell line which is capable of flocculation. These proteins could be proteins implicated in metastasis of cancer, translation initiation factors, proteins involved in hormone signalling pathways such as the insulin or estrogen signalling pathways, proteins involved in mediating cellular adhesion through phosphorylation or known adhesion molecules such as the cadherins and selectins.

In summary this work has expanded our understanding of the biological events at work during the development of a suspension growth phenotype. This knowledge has the potential to be helpful as the biopharma industry moves towards developing novel host cell lines for the expression of novel biopharmaceuticals in the future such as those described by Shen et al. (2019) and Moreira et al. (2020).

5.6. List of Proteins Discussed in This Chapter

HGNC Symbol	Protein Name	Uniprot ID	Listed In
ACACA	Acetyl-CoA carboxylase alpha	Q13085	Table 5.5 and Table 5.6
ACSS2	Acyl-CoA synthetase short chain family member 2	Q9NR19	Table 5.4
ACTB	Actin beta	P60709	Table 5.5 and Table 5.6
AKAP1	A-kinase anchoring protein 1	Q92667	Table 5.3
ALAD	Aminolevulinate dehydratase	P13716	Table 5.4
ALDOA	Aldolase, fructose-bisphosphate A	P04075	Table 5.5 and Table 5.6
ASAH1	N-acylsphingosine amidohydrolase 1	Q13510	Table 5.6
ATP1A1	ATPase Na+/K+ transporting subunit alpha 1	P05023	Table 5.5
ATP1B3	ATPase Na+/K+ transporting subunit beta 3	P54709	Table 5.5
ATP2A2	ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 2	P16615	Table 5.5
ATP2B1	ATPase plasma membrane Ca2+ transporting 1	P20020	Table 5.5
AXL	Tyrosine-protein kinase receptor UFO	P30530	Table 5.4
CACNA2D1	Calcium voltage-gated channel auxiliary subunit alpha2delta 1	P54289	Table 5.5 and Table 5.6
CAMK1	Calcium/calmodulin dependent protein kinase I	Q14012	Table 5.5 and Table 5.6
CAMK2D	Calcium/calmodulin-dependent protein kinase type II subunit delta	Q13557	Table 5.5 and Table 5.6
CD36	CD36 molecule	P16671	Table 5.6
CLIP1	CAP-Gly domain containing linker protein 1	P30622	Table 5.3

HGNC Symbol	Protein Name	Uniprot ID	Listed In
CNN3	Calponin 3	Q15417	Table 5.3
COX6C2	Cytochrome c oxidase subunit 6C2	P11951	Table 5.4
CRK	Adapter molecule crk	P46108	Table 5.5 and Table 5.6
CRNKL1	Crooked neck pre-mRNA splicing factor 1	Q9BZJ0	Table 5.3
CTPS2	CTP synthase 2	Q9NRF8	Table 5.4
CTSD	Cathepsin D	P07339	Table 5.5 and Table 5.6
EDEM3	ER degradation enhancing alpha-mannosidase like protein 3	Q9BZQ6	Table 5.4
EEF2	Eukaryotic translation elongation factor 2	P13639	Table 5.5 and Table 5.6
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1	Q12805	Table 5.4
E1F4E	Eukaryotic translation initiation factor 4E	P06730	Table 5.5 and Table 5.6
EIF4H	Eukaryotic translation initiation factor 4H	Q15056	Table 5.3
ENDOD1	Endonuclease domain containing 1	O94919	Table 5.4
ENO1	Enolase 1	P06733	Table 5.5 and Table 5.6
FAM107B	Protein FAM107B	Q9H098	Table 5.4
FAR1	Fatty acyl-CoA reductase 1	Q8WVX9	Table 5.4
FASN	Fatty acid synthase	P49327	Table 5.5 and Table 5.6
FKBP4	FKBP prolyl isomerase 4	Q02790	Table 5.5 and Table 5.6
FLOT1	Flotillin 1	O75955	Table 5.5 and Table 5.6

HGNC Symbol	Protein Name	Uniprot ID	Listed In
FLOT2	Flotillin 2	Q14254	Table 5.5 and Table 5.6
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	P04406	Table 5.5 and Table 5.6
GNAI2	G protein subunit alpha i2	P04899	Table 5.5 and Table 5.6
GNAI3	Guanine nucleotide-binding protein subunit alpha-13	Q14344	Table 5.5 and Table 5.6
GNAS	Guanine nucleotide-binding protein G(s) subunit alpha isoforms short	P63092	Table 5.5
GNAQ	Guanine nucleotide-binding protein G(q) subunit alpha	P50148	Table 5.6
GPAT3	Glycerol-3-phosphate acyltransferase 3	Q53EU6	Table 5.3
GRB2	Growth factor receptor bound protein 2	P62993	Table 5.2, Table 5.5 and Table 5.6
GSTM7	Glutathione S-transferase Mu 7	Q80W21	Table 5.4
HDAC2	Histone deacetylase 2	Q92769	Table 5.5
HK1	Hexokinase 1	P19367	Table 5.5 and Table 5.6
HMGCR	3-hydroxy-3-methylglutaryl-CoA reductase	P04035	Table 5.2, Table 5.3, Table 5.4, Table
HMOX1	Heme oxygenase 1	P09601	Table 5.4, Table 5.5 and Table 5.6
HSD17B7	Hydroxysteroid 17-beta dehydrogenase 7	P56937	Table 5.4
HSPA2	Heat shock protein family A (HSP70) member 2	P54652	Table 5.2, Table 5.5 and Table 5.6
HSPA8	Heat shock protein family A (Hsp70) member 8	P11142	Table 5.5
HSPB1	Heat shock protein family B (small) member 1	P04792	Table 5.5 and Table 5.6
HSP90AA1	Heat shock protein 90 alpha family class A member 1	P07900	Table 5.5

HGNC Symbol	Protein Name	Uniprot ID	Listed In
HSP90AB1	Heat shock protein 90 alpha family class B member 1	P08238	Table 5.5 and Table 5.6
ITGAV	Integrin subunit alpha V	P06756	Table 5.5
JUP	Junction plakoglobin	P14923	Table 5.4
KDM2A	Lysine demethylase 2A	Q9Y2K7	Table 5.3
KIF20B	Kinesin family member 20B	Q96Q89	Table 5.3
KRT10	Keratin 10	P13645	Table 5.5 and Table 5.6
KRT14	Keratin 14	P02533	Table 5.2, Table 5.5, and Table 5.6
KRT17	Keratin 17	Q04695	Table 5.2, Table 5.5 and Table 5.6
LDHA	Lactate dehydrogenase A	P00338	Table 5.5 and Table 5.6
LGALS1	Galectin 1	P09382	Table 5.3
LGMN	Legumain	Q99538	Table 5.4
LOXL1	Lysyl oxidase like 1	Q08397	Table 5.3
LPL	Lipoprotein lipase	P06858	Table 5.4
LTA4H	Leukotriene A4 hydrolase	P09960	Table 5.4
LUC7L2	LUC7 like pre-mRNA splicing factor	Q9Y383	Table 5.3
MAPK1	Mitogen-activated protein kinase 1	P28482	Table 5.5 and Table 5.6
MAP2K1	Mitogen-activated protein kinase kinase 1	Q02750	Table 5.5
MCL1	Induced myeloid leukaemia cell differentiation protein Mcl-1	Q07820	Table 5.4

HGNC Symbol	Protein Name	Uniprot ID	Listed In
MCM5	Minichromosome maintenance complex component 5	P33992	Table 5.3
MDH1	Malate dehydrogenase 1	P40925	Table 5.4
MMP9	Matrix metallopeptidase 9	P14780	Table 5.2, Table 5.5 and Table 5.6
NCAM1	Neural cell adhesion molecule 1	P13591	Table 5.4
NCOA5	Nuclear receptor coactivator 5	Q9HCD5	Table 5.4
NDUFA5	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5	Q16718	Table 5.5 and Table 5.6
NDUFA9	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial	F5H0J3	Table 5.5
NDUFA10	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial	O95299	Table 5.5 and Table 5.6
NDUFA13	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13	Q9P0J0	Table 5.5 and Table 5.6
NDUFB5	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 5, mitochondrial	O43674	Table 5.5
NDUFB7	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 7	P17568	Table 5.5 and Table 5.6
NDUFB8	Isoform 1 of NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 8,	O95169-1	Table 5.5 and Table 5.6
NDUFB9	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9	Q9Y6M9	Table 5.5 and Table 5.6
NDUFB10	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10	O96000	Table 5.5
NDUFB11	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 11, mitochondrial	Q9NX14	Table 5.5
NDUFS1	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	P28331	Table 5.5 and Table 5.6
NDUFS2	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	O75306	Table 5.5 and Table 5.6
NDUFS5	NADH dehydrogenase [ubiquinone] iron-sulfur protein 5	O43920	Table 5.5 and Table 5.6

HGNC Symbol	Protein Name	Uniprot ID	Listed In
NDUFS8	NADH dehydrogenase [ubiquinone] iron-sulfur protein 8, mitochondrial	O00217	Table 5.5 and Table 5.6
NDUFV1	NADH:ubiquinone oxidoreductase core subunit V1	P49821	Table 5.5
NDUFV2	NADH:ubiquinone oxidoreductase core subunit V2	P19404	Table 5.5 and Table 5.6
NDUFAF1	NADH:ubiquinone oxidoreductase complex assembly factor 1	Q9Y375	Table 5.4
NDUFB8	NADH:ubiquinone oxidoreductase subunit B8	Q95169	Table 5.2
NEU2	Neuramidase 2	Q9Y3R4	Table 5.3
NOLC1	Nuclear and coiled-body phosphoprotein 1	Q14978	Table 5.3
NONO	Non-POU domain containing octamer binding	Q15233	Table 5.3
NT5DC1	5'-nucleotidase domain-containing protein 1	Q5TFE4	Table 5.4
P4HA1	Prolyl 4-hydroxylase subunit alpha-1	P13674	Table 5.3
PAK2	P21 (RAC1) activated kinase 2	Q13177	Table 5.2, Table 5.5 and Table 5.6
PBXIP1	PBX homeobox interacting protein 1	Q96AQ6	Table 5.3
PCK2	Phosphoenolpyruvate carboxykinase 2, mitochondrial	Q16822	Table 5.5 and Table 5.6
PDHA1	Pyruvate dehydrogenase E1 subunit alpha 1	P08559	Table 5.5 and Table 5.6
PDHB	Pyruvate dehydrogenase E1 subunit beta	P11177	Table 5.5
PFKL	Phosphofructokinase, liver type	P17858	Table 5.5 and Table 5.6
PFKP	Phosphofructokinase, platelet	Q01813	Table 5.2 and Table 5.6
PGAM1	Phosphoglycerate mutase 1	P18669	Table 5.5 and Table 5.6

HGNC Symbol	Protein Name	Uniprot ID	Listed In
PGK1	Phosphoglycerate kinase 1	P00558	Table 5.5 and Table 5.6
PKM	Pyruvate kinase M1/2	P14618	Table 5.5 and Table 5.6
PLAUR	Plasminogen activator, urokinase receptor	Q03405	Table 5.4
PLCG1	Phospholipase C gamma 1	P19174	Table 5.2, Table 5.5, Table 5.6
PLIN2	Perilipin 2	Q99541	Table 5.3
PPP1R12A	Protein phosphatase 1 regulatory subunit 12A	O14974	Table 5.2, Table 5.5 and Table 5.6
PPP2R1A	protein phosphatase 2 scaffold subunit A alpha	P30153	Table 5.6
PPP2R2A	protein phosphatase 2 regulatory subunit B alpha	P63151	Table 5.6
PPP3CA	Protein phosphatase 3 regulatory subunit B alpha	P63098	Table 5.5 and Table 5.6
PPP3R1	Protein phosphatase 3 regulatory subunit B alpha	P63098	Table 5.2, Table 5.5 and Table 5.6
PRKAR1A	Protein kinase cAMP-dependent type I regulatory subunit alpha	P10644	Table 5.5 and Table 5.6
PRKAR2B	cAMP-dependent protein kinase type II-beta regulatory subunit	P31323	Table 5.5 and Table 5.6
PTGS2	Prostaglandin-endoperoxide synthase 2	P35354	Table 5.2, Table 5.4, Table 5.5 and
PTPA	Serine/threonine-protein phosphatase 2A activator	Q15257	Table 5.4
PTPN1	Protein tyrosine phosphatase non receptor type 1	P18031	Table 5.3
PTTG1IP	PTTG1 interacting protein	P53801	Table 5.3
PYGB	Glycogen phosphorylase B	P11216	Table 5.5 and Table 5.6
PYGM	Glycogen phosphorylase muscle associated	P11217	Table 5.2 and table 5.6

HGNC Symbol	Protein Name	Uniprot ID	Listed In
RBM3	RNA binding motif protein 3	P98179	Table 5.3
RHOG	Ras homolog family member G	P84095	Table 5.3
RNMT	RNA guanine-7 methyltransferase	O43148	Table 5.4
RRP1B	Ribosomal RNA processing	Q14684	Table 5.3
SCD1	Stearoyl-CoA desaturase	O00767	Table 5.3
SDF2	Stromal cell derived factor 2	Q99470	Table 5.3
SEMA3B	Semaphorin 3B	Q13214	Table 5.4
SLC25A1	Solute carrier family 25 member 1	P53007	Table 5.3
SLC25A5	solute carrier family 25 member 5	P05141	Table 5.5
SLC2A1	Solute carrier family 2 member 1	P11166	Table 5.3, Table 5.5 and Table 5.6
SLC6A6	Solute carrier family 6 member 6	P31641	Table 5.3
SNX2	sorting nexin 2	O60749	Table 5.4
SQLE	Squalene epoxidase	Q14534	Table 5.3
SPTLC1	Serine palmitoyltransferase 1	O15269	Table 5.6
SRC	SRC proto-oncogene, non-receptor tyrosine kinase	P12931	Table 5.5
STT3B	STT3 oligosaccharyltransferase complex catalytic subunit B	Q8TCJ2	Table 5.3
TALDO1	transaldolase 1	P37837	Table 5.4
THBD	Thrombomodulin	P07204	Table 5.3 and Table 5.4

HGNC Symbol	Protein Name	Uniprot ID	Listed In
TFRC	transferrin receptor	P02786	Table 5.5 and Table 5.6
TMED5	Transmembrane emp24 domain-containing protein 5	Q9Y3A6	Table 5.4
TMEM129	E3 ubiquitin-protein ligase TM129	A0AV14	Table 5.4
TMEM132A	Transmembrane protein 132A	Q24JP5	Table 5.3
TMTC3	Transmembrane O-mannosyltransferase targeting cadherins 3	Q6ZXV5	Table 5.3
TSN	Translin	Q15631	Table 5.3
UHRF1	Ubiquitin like with PHD and ring finger domains 1	Q96T88	Table 5.3
UTP20	UTP20 small subunit processome component	O75691	Table 5.3
VDAC1	voltage dependent anion channel 1	P21796	Table 5.5
VDAC2	voltage dependent anion channel 2	P45880	Table 5.5
VDAC3	voltage dependent anion channel 3	Q9Y277	Table 5.5
VIM	Vimentin	P08670	Table 5.3
VPS29	Vascular protein sorting-associated protein 29	Q05DG7	Table 5.4
XPNPEP1	Xaa-Pro aminopeptidase 1	Q9NQW7	Table 5.4
YARS	Tyrosine--tRNA ligase, cytoplasmic	P54577	Table 5.4

Table 5.9: **Protein names associated with the HGNC symbols listed in Table 5.2, Table 5.3, Table 5.4, Table 5.5, and Table 5.6.**

CHAPTER 6

A proteomic comparison of CHO K1 and DXB11 cells following adaption from attached to suspension growth in the presence of FBS

6.1. Introduction

One of the reasons why CHO cells became the predominant cell line used in the biopharmaceutical industry was the development of CHO cell lines capable of suspension growth in serum free chemically defined media. This allowed the development of high titre, economically viable processes at scales of up to 20,000L in fed batch stirred tank reactors. Murata et al. (1988) and Boraston et al. (1992) reported the successful adaptation of CHO K1 cells to suspension growth by a serial dilution method and manipulating the growth media, respectively. In 1996 Sinacore et al. reported on the generation of DUXB11 cells which were preadapted for suspension growth. Prior to this advance each new cell line construction project involved a period where cells had to be adapted to suspension growth. These studies focused more on the topic of how to promote suspension growth rather than answering the question of what was changing in the cell's proteome and genome. Walther et al. (2016) compared the cell surface proteome of an adherent (CCL-61) and a proprietary suspension CHO cell line by flow cytometry and confocal microscopy. They proposed that the integrin beta 1 participated in an "inside out" signalling event which was central to the adaption of CHO cells to suspension growth. Furthermore, they demonstrated that the transition, of the two cell lines being studied, from adherent to suspension growth correlated with the downregulation of tyrosine 397 phosphorylation in the enzyme focal adhesion kinase (FAK). The role FAK plays in cell migration is well understood (Sieg et al.; 1999, Zhao and Guan; 2011) and it was established by Cary et al. (1996) that the phosphorylation of tyrosine 397 was necessary before FAK could promote cell migration in CHO cells. Finally, Yu et al. (2018) demonstrated that the inhibition of FAK correlated with decreased cellular adhesion due to reduced numbers of filopodia and changes in the localisation of actin fibres.

Previous chapters discussed the challenges associated with the clarification of high-density cell lines in industrial settings. Potential solutions which were investigated included:

- a cell engineering solution in which four eukaryotic proteins, all of which were exogenous to CHO cells, were identified following a literature review and cloned

in CHO cells in an attempt to develop a flocculating CHO cell line. However, none of the proteins investigated in that study appeared to be expressed in CHO cells.

- a proteomic approach to identifying endogenous proteins that might play a role in cell adhesion using label free differential expression analysis by comparing the proteome of CHO K1 cells grown attached to a surface or in suspension. However, as previously mentioned in chapter 5, that study used CHO K1 cells which were preadapted to suspension culture.

Consequently, it was deemed appropriate to build on that study using cells which hadn't been adapted to either suspension or growth in serum free media before. Besides identifying potential engineering targets, using adapted cells would provide the opportunity to better understand the pathways and processes impacted by or underpinning the adaptation process more generally – both from attached to suspension culture but also from serum-dependence to serum-free growth.

The adaption of CHO cells from attached to suspension culture is described in this chapter (chapter 6) while the adaptation of CHO cells from serum-dependence to serum-free growth is described in chapter 7. This latter adaptation (to serum independence) has also been anecdotally associated with a tendency for cells to aggregate or clump thus keeping with the overall aims of this thesis.

To avoid the use of preadapted CHO K1 cells, a vial of attached CHO K1 cells was purchased from the European Collection of Authenticated Cell Culture (ECACC). These cells were initially cultured in 5% serum and then adapted to suspension growth. Additionally, a vial of CHO DXB11 cells was obtained from Prof. Larry Chasin of Columbia University in New York as a second comparison. These were similarly cultured in 5% serum and then adapted to suspension growth.

6.2. Experimental Design and Sample Preparation

To identify proteins, which differed in their expression levels between the experimental conditions under investigation, a comparative analysis was performed using a label-free

quantitative differential expression LC-MS approach. The approach taken utilised four biological replicates of each condition as outlined in Figure 6.1.

Non-producing parental CHO-K1 and CHO-DXB11 cell lines were used for this experiment. The CHO-K1 cells were revived from a frozen vial and grown for three passages in T75 flasks in Balan CD growth medium A supplemented with 5% FBS and 4mM L-Glutamine. The CHO-DXB11 cells were donated by Prof L Chasin as a viable culture. These were transferred into αMEM containing Hypoxanthine, Thymidine, 5% FBS and 4mM L-Glutamine. All T75 flasks were cultured in a static incubator at 37°C, 5% CO_2, 80% RH. Once cells were ~80% confluent, they were trypsinised, and new flasks were seeded at $2.5X10^5$ cells/flask (Section 2.2.1). After three passages, the cells were established as attached and suspension cultures as follows:

- K1 cells were transferred to four T175 flasks, seeded at $1.0X10^6$ cells/flask. The media used was as described above (K1 Attached).
- K1 cells were transferred to four 250mL Corning® shake flasks (30mL working volume), seeded at $2.5X10^5$ cells/mL, in a Kühner orbital shaker at 37°C, 5% CO_2, 80% RH and 130 rpm (K1 suspension). The media used was identical to that described for the K1 attached cells.
- DXB11 cells were transferred to four T175 flasks, seeded at $1.0X10^6$ cells/flask. The media used was αMEM containing Hypoxanthine, Thymidine, 5% FBS and 4mM L-Glutamine. (DXB11 Attached 1).
- DXB11 cells were transferred to four T175 flasks, seeded at $1.0X10^6$ cells/flask (DXB11 Attached 2). The media used was Balan CD growth medium A, supplemented with Hypoxanthine, Thymidine, 5% FBS and 4mM L-Glutamine.
- DXB11 cells were transferred to four 250mL Corning® shake flasks (30mL working volume), seeded at $2.5X10^5$ cells/mL, in a Kühner orbital shaker at 37°C, 5% CO_2, 80% RH and 130 rpm. (DXB11 suspension 1). The media was as described above for the DXB11 attached 1 cells.
- DXB11 cells were transferred to four 250mL Corning® shake flasks (30mL working volume), seeded at $2.5X10^5$ cells/mL, in a Kühner orbital shaker at 37°C, 5% CO_2, 80% RH and 130 rpm. (DXB11 suspension 2). The media used was Balan CD

growth medium A supplemented with Hypoxanthine, Thymidine, 5% FBS and 4mM L-Glutamine.

After a further passage the attached cells were harvested by scraping, the cells pelleted by centrifugation, washed with ice cold PBS, and stored at -80°C until required. Suspension cultures were counted daily and passaged every three days until culture viability was above 90% and cells were doubling approximately every day. At this point cells were considered adapted to suspension growth. The cells were pelleted by centrifugation, washed with ice cold PBS, and stored at -80°C until required.

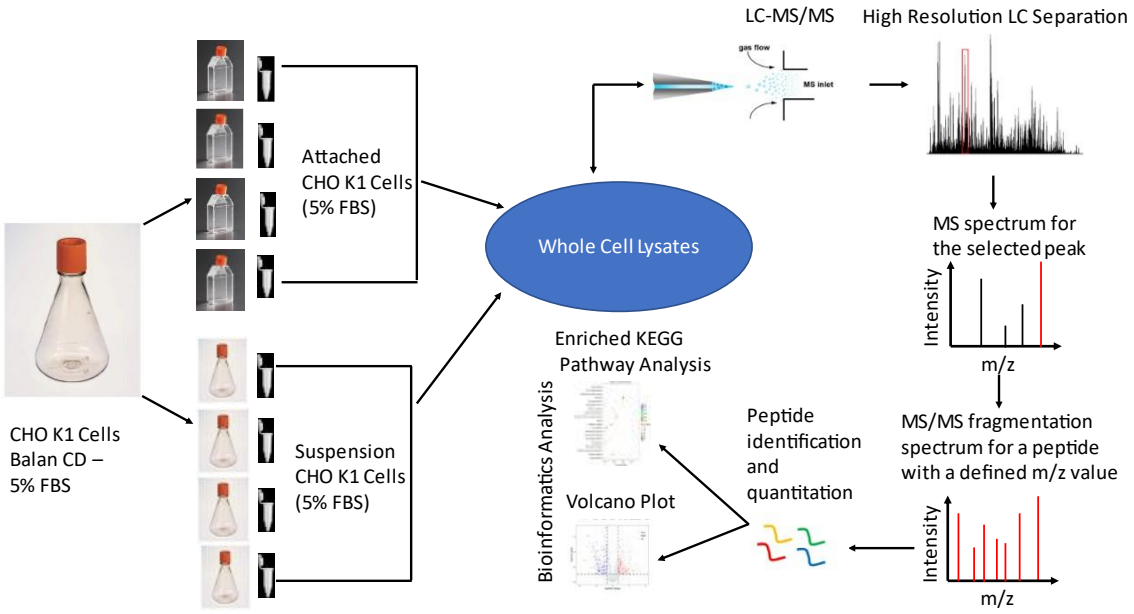


Figure 6.1: **Workflow followed when preparing CHO K1 samples, from cells which had been adapted from attached to suspension growth in media containing serum, for LC-MS analysis. The same approach was followed for the DXB11 cells which had undergone the same adaption process.**

Ideally, all cells should be in the same media to simplify the comparison. The purpose of setting up DXB11 suspension cultures in both αMEM and Balan CD was to have a backup should it not be possible to establish a suspension culture in either medium.

The sample preparation for the proteomics experiment consisted of the following steps:

1. Lysis of cell pellets to release proteins.
2. Quantification/normalisation of released proteins between samples.
3. Denaturation, alkylation, and digestion of the proteins to produce peptides.
4. Resuspension of the peptides in a mass spec compatible buffer.

In this project these steps were achieved utilising the FASP protocol developed by Coleman et al. (2017) (Section 2.6.3.7). To reduce variability and retention time drift, all timepoints corresponding to each cell line were prepared and analysed on the mass spectrophotometer in a single analysis and the "Two-Sample Analysis Approach'' described by Kelly (2016) was adopted.

Following alignment, samples were split into experimental groups and identified peptides were filtered based on an ANOVA p value of ≤ 0.05 between experimental groups. A mascot generic file (.mgf) was generated for each replicate and exported to Proteome Discoverer. Following peptide and protein identification by Proteome Discoverer, using a proteogenomic draft annotation of the CHO genome Li et al. (2018) (Section 2.6.5), the data was saved as a (.pepxml) file and reimported into Progenesis QI for Proteomics. Criteria of an ANOVA p value of ≤ 0.05, a relative minimum fold change of +/-1.5 between experimental groups and ≥2 peptides contributing to protein identifications was used to identify proteins whose expression profile differed between experimental groups.

All subsequent data manipulation, protein naming, and statistical analysis was performed in the R programming environment (Section 2.6.6).

6.3. Results

Cell pellets corresponding to attached and suspension adapted cultures were processed for MS analysis as per Sections 2.6.4 and 2.6.5. The resulting DE protein lists were subjected to:

a) Volcano plots of the DE proteins detected in each experimental condition were plotted using the Enhanced Volcano package in R (Blige et al., 2021).
b) Upset plots were used to identify the number of DE proteins which were unique or common to CHO K1 and CHO DXB11 cells lines. The R package UpSetR (Conway et al., 2017) was used to draw these.
c) Peptide set enrichment analysis (PSEA) of the data, using the pathfindR package in R (Ulgen et al., 2019), to identify statistically enriched KEGG pathways ($p \leq 0.05$) and whether the proteins identified in those pathways were upregulated or downregulated.
d) Identification of enriched Gene Ontology (GO) terms ($p \leq 0.05$), related to the DE proteins identified by PSEA, using the enrichGO function from the Annotation Hub R package (Morgan and Shepard, 2021).
e) Categorisation of enriched GO terms, with respect to the cellular biological process they described, using the rrvgo R package (Sayols, 2020)

6.3.1. Assessment of replicate clusters

Two approaches were taken to the clustering analysis. These were principal component analysis (PCA) which is part of the Progenesis QI for Proteomics workflow (Section 2.6.8), and the use of heat maps (Section 2.6.7) based on the Euclidean distance.

PCA is a statistical technique which allows the description of the maximum amount of information in a dataset using the least possible number of variables (known as principal components). It has as its aim, increasing ease of interpretation while minimising information loss. In an ideal scenario all the variation in a dataset would be described by 1 variable (principal component 1) but this rarely (if ever) occurs.

Performing PCA on the 5% attached and 5% suspension cells (Figure 6.2) revealed that the percentage variability described by PCA one and two was 77% with most of the variability (68%) being described by principal component one. The variability among the attached cells is described by both principal component 1 and 2 while the variability among suspension cells is described by principal component one only. Analysis of Figure 6.2 supports the theory that adaption of CHO K1 cells to suspension growth results in distinct proteomic changes as CHO K1 cells grown as suspension cultures cluster away

from CHO K1 cells grown as attached cultures and two distinct groups of DE proteins can be observed.

The PCA analysis of DXB11 cells (Figure 6.3) revealed that the percentage variability described by PCA one and two was 74%, during the adaptation from attached to suspension growth, with most of the variability (65%) being described by principal component one. The variability among the attached cells is described by both principal component 1 and 2 while the variability among suspension cells is described by principal component two only. Analysis of Figure 6.3 supports the theory that adaption of CHO DXB11 cells to growth in serum free media results in distinct proteomic changes as CHO DXB11 cells grown in the absence of serum cluster away from CHO DXB11 cells grown in the presence of serum as two distinct groups of DE proteins can be observed.

The greater spread observed with the attached CHO DXB11 cells may be explained by the fact that this experimental condition encompasses four replicate cultures which were samples by scrapping the cells off the flask. Alternatively, as discussed in chapter five it could be due to variability in sample preparation and the total number of proteins detected in each replicate. To investigate this, the total number of proteins detected in each replicate was quantified using Proteome Discoverer. These numbers are displayed in Table 6.1.

As an additional quality control step heatmaps (Figure 6.4 and Figure 6.5) based on the Euclidean distance were also used to examine the clustering of the samples. All four replicates from each experimental group (attached cells 5% FBS and, suspension cells 5% FBS) were analysed.

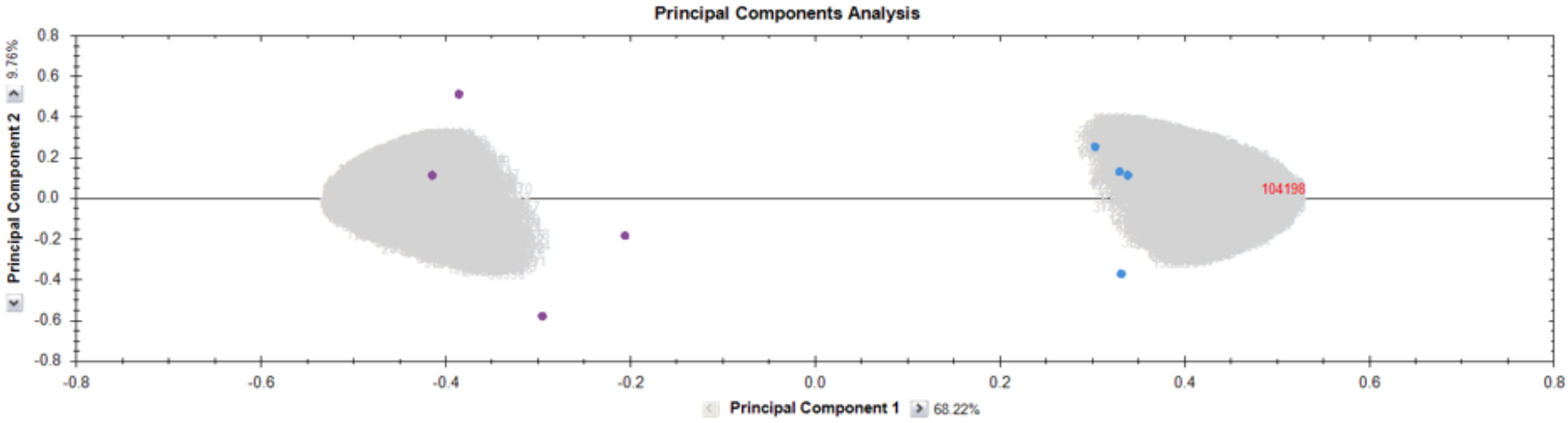


Figure 6.2: **Principal Component Analysis displaying the clustering pattern of CHO K1 cells which have adapted from attached (5% serum) to suspension (5% serum) growth (N=4). PCA analysis was performed by the Progenesis QI for Proteomics software after identification of DE proteins as a quality control step to assess the clustering of replicate samples.** *Blue dots represent suspension adapted cells grown in media containing 5% serum. Purple dots represent attached cells grown in media 5% serum. The grey clouds represent individual proteins found to be differentially expressed in each condition.*

The greater spread observed between replicates of the attached CHO K1 cells might be explained by the fact that this experimental condition encompasses four replicate cultures which were sampled by scrapping the cells off the surface of the flask.

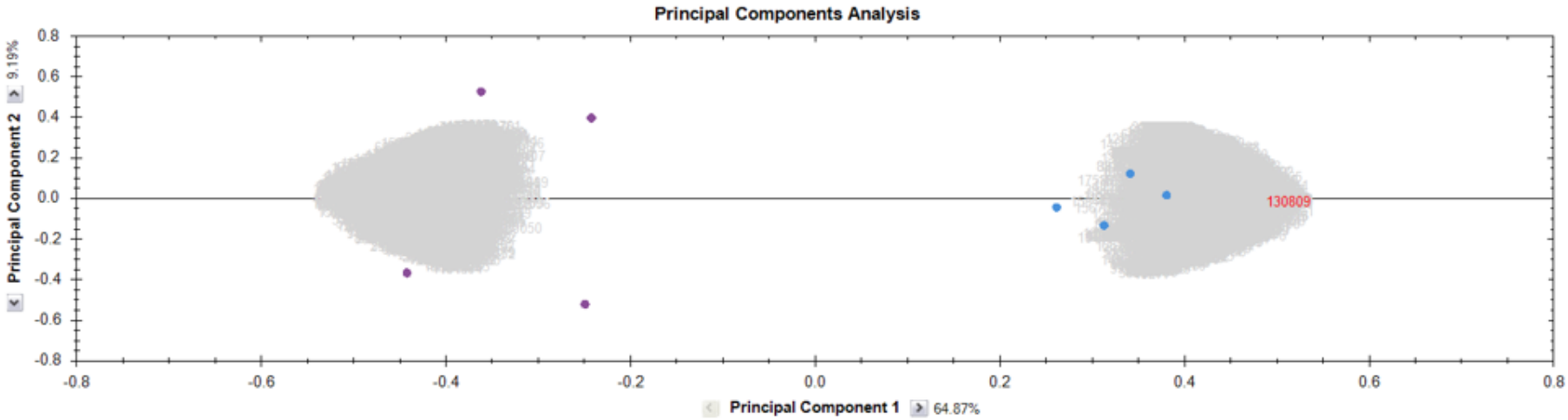


Figure 6.3: **Principal Component Analysis displaying the clustering pattern of CHO DXB11 cells which have adapted from attached (5% FBS) to suspension (5% FBS) growth (N=4). PCA analysis was performed by the Progenesis QI for Proteomics software after identification of DE proteins as a quality control step to assess the clustering of replicate samples.** *Blue dots represent suspension adapted cells grown in media containing 5% serum. Purple dots represent attached cells grown in media containing 5% serum. The grey clouds represent individual proteins found to be differentially expressed in each condition.*

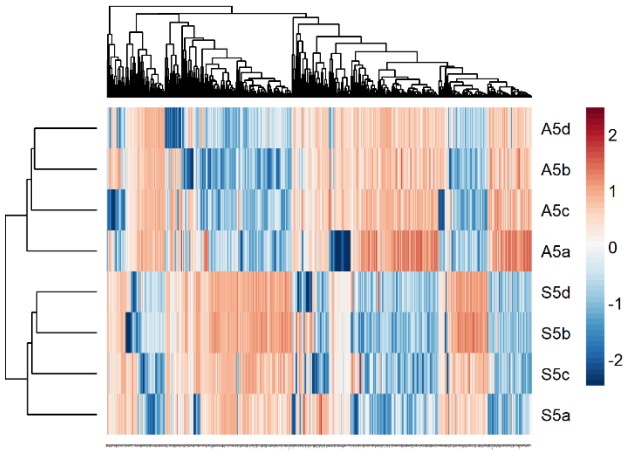


Figure 6.4: **Clustering of replicates based on the Euclidean distance between samples. All CHO K1 DE proteins with a p ≤ 0.05 and fold change ≥ 1.5 (identified by Progenesis QI for Proteomics) are displayed.**

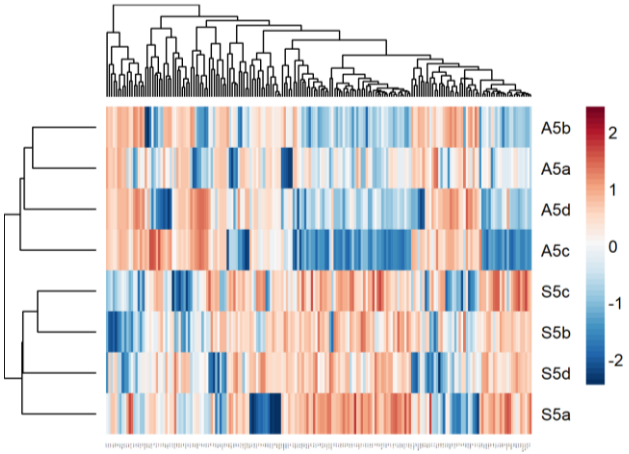


Figure 6.5: **Clustering of replicates based on the Euclidean distance between samples. All CHO DXB11 DE proteins with a p ≤ 0.05 and fold change ≥ 1.5 (identified by Progenesis QI for Proteomics) are displayed.**

Figures 6.4 and 6.5 confirm that that replicates from each experiment group cluster together.

Experimental Group	Replicate	Total No. of Proteins	Average No. of Proteins	No. of Common Proteins	No. of DE Proteins
K1 Attached (5% FBS)	A5a	3070	3181	2336	FC = 1.5
	A5b	3089			
	A5c	3359			
	A5d	3207			
K1 Suspension (5% FBS)	S5a	3261	3323	2471	
	S5b	3480			
	S5c	3043			
	S5d	3506			
DXB11 Attached (5% FBS)	A5a	3510	3658	2598	FC = 1.5
	A5b	3301			
	A5c	3744			
	A5d	4077			
DXB11 Suspension (5% FBS)	S5a	4157	3238	1714	
	S5b	4082			
	S5c	1967			
	S5d	2745			

Table 6.1: **Total numbers of proteins detected in all replicates. MS raw data files were imported directly into Proteome Discoverer and proteins detected as per Section 2.6.5. The number of common proteins column recorded the number of proteins detected in all four replicates.**

In excess of three thousand proteins were identified in each replicate for both K1 experimental groups and the DXB11 5% FBS attached experimental group. However, in the DXB11 5% suspension experimental group greater variability was observed between replicates with the number of proteins detected ranging from one thousand nine hundred and sixty seven to four thousand one hundred and fifty seven. The impact of this variability is that not all proteins were detected in each replicate. In order to assess the impact of this the proteins common to all replicates of each experimental group was identified and was compared to the proteins identified as playing a role in the KEGG pathways (Section 6.3.3). This enabled the identification of any proteins present in these pathways which were not common to all replicates. These proteins are listed in Table 6.2.

KEGG Path	HGNC Symbol	Protein Name	UniProt ID	Direction
Tight Junction	NEDD4	E3 ubiquitin protein ligase NEDD4	P46934	Down Regulated
Endocytosis	NEDD4	E3 ubiquitin protein ligase NEDD4	P46934	Down regulated
Alanine, aspartate and glutamate metabolism	ALDH5A1	aldehyde dehydrogenase 5 family member A1	P51649	Up Regulated

Table 6.2: **DE proteins detected in enriched KEGG paths which were not detected all eight replicates of each condition.**

Only two proteins were identified as part of enriched KEGG pathways which were not present in all sample replicates. *E3 ubiquitin protein ligase* (*NEDD4*) was found to be present in all four replicates of the DXB11 5% FBS attached cells and was only present in one replicate of the DXB11 5% FBS suspension cells. It was also detected in all replicates of the K1 5% FBS attached and suspension cells.

Similarly, *Aldehyde dehydrogenase 5 family member A1* (*ALDH5A1*), was found to be present in three of the four replicates for DXB11 5% FBS attached cells and suspension cells. It was also detected in three of four replicates for K1 5% FBS attached cells and all replicates for K1 5% FBS suspension cells.

It was therefore considered that NEDD4 and ALDH5A1 were both true differentially expressed proteins.

6.3.2. Identification of differentially expressed proteins with a common expression profile in CHO K1 and DXB11 cells after adaptation to suspension growth

One thousand six hundred and seventy-eight DE proteins were detected in CHO DXB11 cells compared to one thousand two hundred and eighty-five in CHO K1 cells following adaptation to suspension growth in serum free media. Once the appropriate filters had been applied, max fold change as calculated by Progenesis QI for Proteomics ≥1.5, an p value ≤ 0.05, a false discovery rate of 1% and at least two unique peptides contributing

to protein identification, one thousand and forty-four of these DE proteins were identified as being unique to CHO K1 cells. (Appendix 9). The same filtering step identified one hundred and ninety DE proteins as unique to CHO DXB11 cells (Appendix 10).

Table 6.3 lists the top five proteins in either cell line based on:

a. Absolute fold change (absolute FC) between the two experimental groups
b. p value

Cell Line	Protein Name	HGNC Symbol	p Value	Max FC[1]	Log_2 (Absolute FC)[2]	Present in all attached reps	Present in all suspension reps
Ranked by absolute FC							
K1	Angiopoietin 4	ANGPT4	$5.49E10^{-4}$	100.7	6.6	Yes	Yes
	Prothymosin alpha 1	PTMA	$3.62E10^{-2}$	72.4	-6.1	Yes	Yes
	Glycerol-3-phosphate dehydrogenase	GPD1	$2.57E10^{-4}$	20.0	4.3	Yes	Yes
	S100 calcium binding protein	S100A4	$1.32E10^{-5}$	14.1	3.8	Yes	Yes
	Pigment epithelium-derived factor	SERPINF1	$4.50E10^{-5}$	12.9	3.6	Yes	Yes
DXB11	T-complex protein 1 subunit theta	CCT8	$3.12E10^{-2}$	43.2	5.4	Yes	Yes
	3'-phosphoadenosine 5'-phosphosulfate synthase 1	PAPSS1	$1.33E10^{-2}$	21.9	4.5	Yes	Yes
	Prolyl 3-Hydroxylase 1	P3H1	$2.32E10^{-2}$	16.2	4.0	Yes	Yes

Cell Line	Protein Name	HGNC Symbol	p Value	Max FC[1]	Log_2 (Absolute FC)[2]	Present in all attached reps	Present in all suspension reps
DXB11	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1 like	MTHFD1L	$1.85E10^{-3}$	13.1	3.7	Yes	Yes
	E3 ubiquitin protein ligase NEDD4	NEDD4	$1.64E10^{-2}$	0.1	-3.4	Yes	Only present in replicate S5a
Ranked by p value (order of significance)							
K1	Glutathione S-transferase pi 1	GSTP1	$5.97E10^{-8}$	2.4	1.2	Yes	Yes
	Anilin actin binding protein	ANLN	$6.14E10^{-8}$	4.2	-2.1	Yes	Yes
	Uridine-cytidine kinase 2	UCK2	$1.50E10^{-8}$	5.0	-2.3	Yes	Yes
	Acetyl-CoA acyltransferase 2	ACAA2	$3.49E10^{-8}$	3.3	1.7	Yes	Yes
	Transglutaminase 2	TGM2	$6.39E10^{-8}$	3.1	1.6	Yes	Yes
DXB11	Annexin A2	ANXA2	$4.85E10^{-2}$	1.7	0.7	Yes	Yes

Cell Line	Protein Name	HGNC Symbol	p Value	Max FC[1]	Log_2 (Absolute FC)[2]	Present in all attached reps	Present in all suspension reps
DXB11	Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2	$4.66E10^{-2}$	5.2	2.4	Yes	Yes
	T-complex protein 1 subunit delta	CCT4	$4.60E10^{-2}$	1.6	-0.7	Yes	Yes
	Tricarboxylate transport protein, mitochondrial	SLC25A1	$4.58E10^{-2}$	1.8	0.9	Yes	Yes
	Karyopherin subunit beta 1	KPNB1	$4.56E10^{-2}$	1.5	-0.6	Yes	Yes

[1] Max FC is the proteins fold change value as calculated by Progenesis QI for Proteomics.

[2] Log2 Absolute FC allows a direction to be placed on the fold change. A value > 0 indicated the protein is upregulated in the comparison while a value < 0 indicates the protein is downregulated. It was calculated by dividing the proteins mean abundance in suspension cells by its mean abundance in attached cells.

Table 6.3: The five most significant DE proteins detected in CHO K1 and CHO DXB11 cell lines as ranked by max fold change and statistical significance.

While the subsequent analysis could be confined to these proteins a different approach was taken. It was decided to focus on DE proteins which were common to both cell lines and displayed a common expression profile in response to the adaptation to suspension Bearing in mind what the experiment was designed to identify - proteins implicated in the adaptation of cells from attached to suspension growth – it seemed logical to focus on these proteins rather than those with the lowest p value or highest fold change.

Two different techniques (volcano plots and upset plots) were used to examine all DE proteins detected in both K1 and DXB11 cells visually.

Volcano plots of the DE proteins detected in CHO K1 and CHO DXB11 cells is presented in Figures 6.6 and 6.7. Volcano plots of statistical significance ($-\log_{10}$Pvalue) versus fold change ($\log_2$ FC) provide an overall impression of whether and how many proteins are up or down regulated within a dataset. By $\log_2$ transforming the data before plotting the plot can be centred around zero. DE proteins with a Log_2 FC > 0 are considered upregulated (these fall on the right hand side of Figures 6.6 and 6.7. DE proteins on the left-hand side of Figures 6.6 and 6.7 have a Log_2 FC < 0 and are downregulated. Similarly, by log transforming the p values associated with each data point ($-\text{Log}_{10}$) highly significant (v small p values) proteins are positioned towards the top of the plot and while those clustered towards the bottom of the plot are less significant.

The fold change in an individual proteins' expression was defined as

$$\frac{average\ normalised\ protein\ abundance\ in\ cells\ adapted\ to\ suspension\ growth}{average\ normalised\ protein\ abundance\ in\ attached\ cells}$$

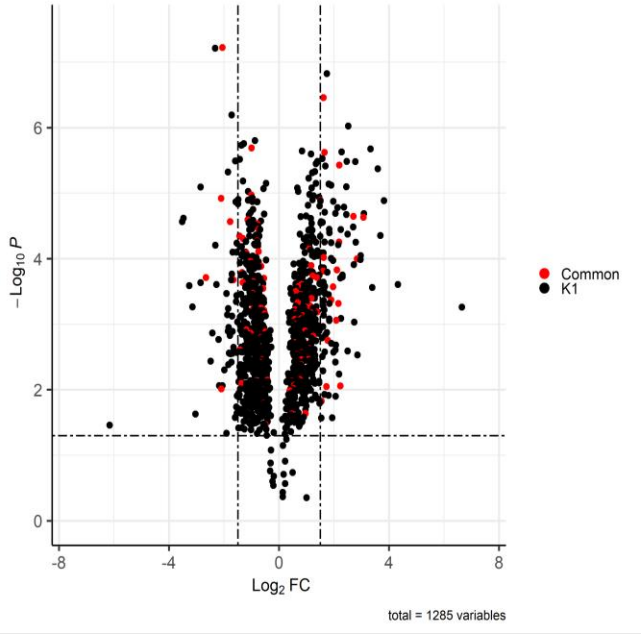


Figure 6.6: All one thousand two hundred and eighty-five DE proteins detected in CHO K1 cells following adaptation from attached to suspension growth in the presence of serum (5% FBS). Red dots represent the one hundred DE proteins common to both CHO K1 and DXB11 cells. The vertical dashed lines in Figure 6.7 represent a fold change of ± 1.5 and the horizontal dashed line represents $\log_{10}(0.05)$.

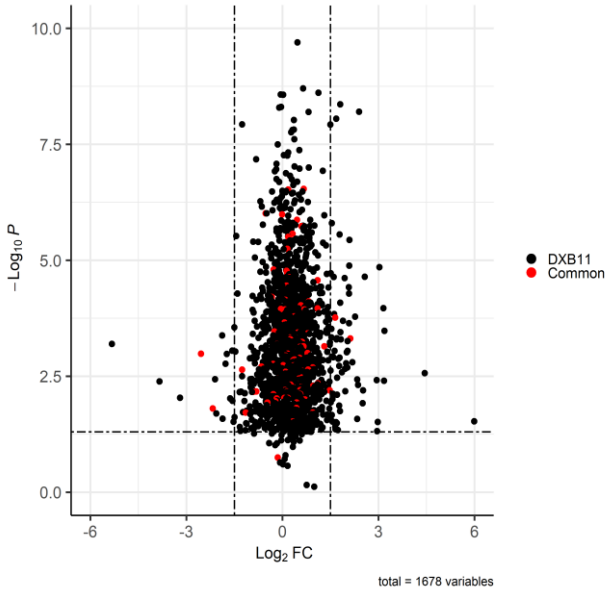


Figure 6.7: All one thousand six hundred and seventy-eight DE proteins detected in CHO DXB11 cells following adaptation from attached to suspension growth in the presence of serum (5% FBS). Red dots represent the one hundred DE proteins common to both CHO K1 and DXB11 cells. The vertical dashed lines in Figure 6.8 represent a fold change of ± 1.5 and the horizontal dashed line represents $\log_{10}(0.05)$.

Upset plots, Lex et al. (2014) (Figures 6.8 and 6.9) were used to identify common DE proteins and gene ontology terms between different subpopulations of DE proteins detected in the two cells lines. Upset plots are a variation of a Venn diagram, with the advantage that they allow an increased number of subpopulations to be examined in a single plot compared to Venn diagrams. The subpopulations under comparison were:

a) Proteins whose abundance was higher in suspension cells than attached cells in the DXB11 cell line. These proteins were classified as part of the Up.DXB11 subset.
b) Proteins whose abundance was lower in suspension cells than attached cells in the DXB11 cell line. These proteins were classified as part of the Down.DXB11 subset.
c) Proteins whose abundance was higher in suspension cells than attached cells in the K1 cell line. These proteins were classified as part of the Up.K1 subset.
d) Proteins whose abundance was lower in suspension cells than attached cells in the K1 cell line. These proteins were classified as part of the Down.K1 subset.

The size of each subpopulation (Set Size) is represented by the small bar chart at the bottom of Figures 6.8 and 6.9. The overlap between each subpopulation (Intersection Size) is represented by the y-axis of main bar chat. What each intersection is, is explained by the grid on the x-axis of the main bar chart.

The maximum number of sub populations/intersections four distinct populations can have is fifteen. There's a greater degree of overlap between subpopulations in Figure 6.8 compared to Figure 6.9. This is because each GO term will have multiple proteins associated with it and it's possible that a protein which is upregulated in one cell line and downregulated in another will share a common GO term. For example, the GO term GO:2000049 (positive regulation of cell-cell adhesion mediated by cadherin) is common to both Flot1 which is upregulated in K1 and TJP1 which is downregulated in DXB11. Therefore, no further analysis was performed on Figure 6.8 and instead GO analysis was performed following peptide set enrichment analysis (PSEA).

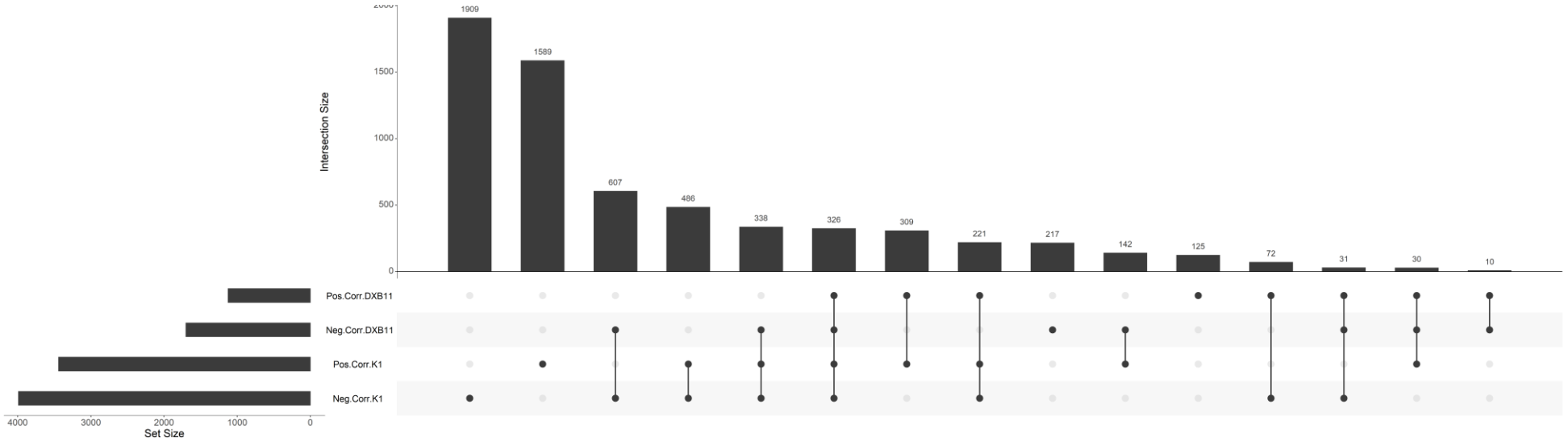


Figure 6.8: **Number of unique Gene Ontology terms present in each of the subpopulations of interest and the number of GO terms which overlap between two or more subpopulations. The numbers above the bars refer to the number of DE proteins represented by that bar. Set Size on the lower left-hand corner of the figure refers to the total number of DE proteins in each subpopulation (e.g., there are 3990 GO terms in total in the Neg.Corr.K1 subpopulation and 1701 GO terms in total in the Neg.Corr.DXB11 subpopulation. However, these two subpopulations have 607 GO terms in common).**

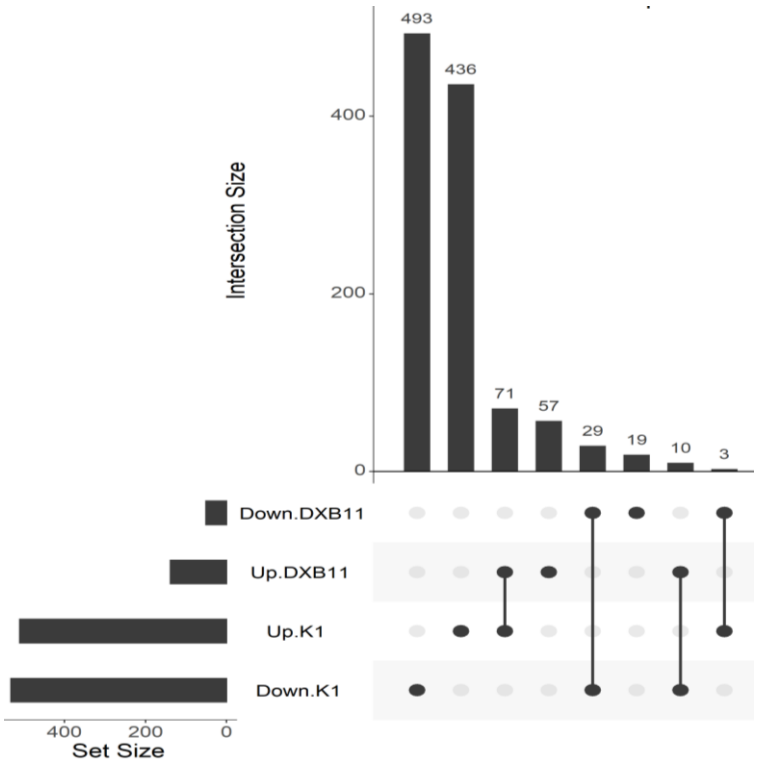


Figure 6.9: **Number of unique DE proteins present in each of the subpopulations of interest and the number of DE proteins which are common to two or more subpopulations. The numbers above the bars refer to the number of DE proteins represented by that bar. Set Size on the lower left-hand corner of the figure refers to the total number of DE proteins in each subpopulation (e.g., there are 493 DE proteins which are unique to the Down.K1 subpopulation and 19 DE proteins which are unique to the Down.DXB11 subpopulation. However, these two subpopulations have 29 DE proteins in common. Sub populations which have no DE proteins in common e.g., Up.DXB11 and Down.DXB11 are not shown in Figure 6.9.**

However, all the differentially expressed proteins presented in Figure 6.9 fell into one of eight subpopulations as described below. This makes sense as it's not possible for the same differentially expressed protein to be a member of both the upregulated DXB11 and downregulated DXB11 sub populations.

- Four hundred and ninety-three proteins were present at lower levels in suspension K1 cells compared to attached cells (Down.K1). In this study they were found to be uniquely DE in CHO K1 cells.
- Four hundred and thirty-six proteins were up regulated in suspension K1 cells compared to attached cells (Up.K1). In this study they were found to be uniquely DE in CHO K1 cells.

- Nineteen proteins whose expression was lower in suspension cells than attached cells (Down.DXB11) were found to be uniquely DE in CHO DXB11 cells.
- Fifty-seven proteins whose expression was higher in suspension cells than attached cells (Up.DXB11) were found to be uniquely DE in CHO DXB11 cells.
- Three DE proteins present in the Up.K1 subpopulation were also found to be present in the Down.DXB11 subpopulation, i.e., their expression profile differs between the two cell lines.
- Ten DE proteins present in the Down.K1 subpopulation were also found to be present in the Up.DXB11 subpopulation, i.e., their expression profile differs between the two cell lines.
- Seventy-one DE proteins common to both CHO K1 and DXB11 cells had a higher expression level in suspension than attached.
- Twenty-nine DE proteins common to both CHO K1 and DXB11 cells had a lower expression level in suspension than attached.

The last two categories comprise the proteins of interest. These are the one hundred DE proteins common to K1 and DXB11 cell lines which exhibited the same expression pattern upon adaptation from attached to suspension growth in the presence of 5% FBS. (Appendix 11)

6.3.3. Identification of common enriched KEGG pathways identified in CHO K1 and DXB11 cells

The rest of the analysis was focused upon the one hundred DE proteins which were common to both cell lines and displayed a common expression profile. Bearing in mind what the experiment was designed to identify - proteins implicated in the adaptation of cells from attached growth to suspension growth – it seemed logical to focus on this set of proteins.

Gene set enrichment analysis (GSEA) is a common technique for identifying groups of genes or proteins which are overrepresented in large data sets. By comparing two conditions and analysing where in the dataset a set of genes or proteins fall inferences can be drawn about whether those genes or proteins are related to phenotypic

differences. PSEA, a variation of GSEA, was developed specifically for the analysis of proteomic datasets. This study used the R package PathfindR (Ulgen et al., 2019) to perform PSEA on the 100 DE proteins identified as being common to both K1 and DXB11 cells under the conditions being tested. The analysis was performed separately on each cell line and the results combined. This approach was taken because although the proteins are common to both K1 and DXB11 cell lines the fold change and p value for each protein differed between the two cells lines. The output of the analysis was a list of statistically enriched KEGG pathways, the DE proteins detected in those paths and whether these proteins were upregulated or downregulated. The output from each pathfindR analysis were compared to see if the enriched KEGG path IDs differed between K1 and DXB11 cell lines. Seventy four KEGG paths were identified as being enriched, of which nine were unique to DXB11, ten were unique to K1 and the remaining fifty five were common to both cell lines (Appendix 12). Additionally, only sixty of the one hundred common DE proteins were identified as being components of these seventy four enriched KEGG paths.

The KEGG system contains, among other information, a collection of pathway maps representing our knowledge of the molecular interaction, reaction, and relation networks for:

- Metabolism
- Genetic Information Processing
- Environmental Information Processing
- Cellular Processes
- Organismal Systems
- Human Diseases
- Drug Development

Each of these six categories is further sub divided into common themes. Therefore, by identifying the individual KEGG pathways which are statistically enriched in the comparison an overall picture of what changed within the cell can be obtained.

Figures 6.10 and 6.11 show the distribution of the seventy four significantly enriched KEGG pathways detected following the adaption of K1 and DXB11 cell lines from

attached to suspension growth across the seven themes which comprise the KEGG system and within the metabolic theme respectively.

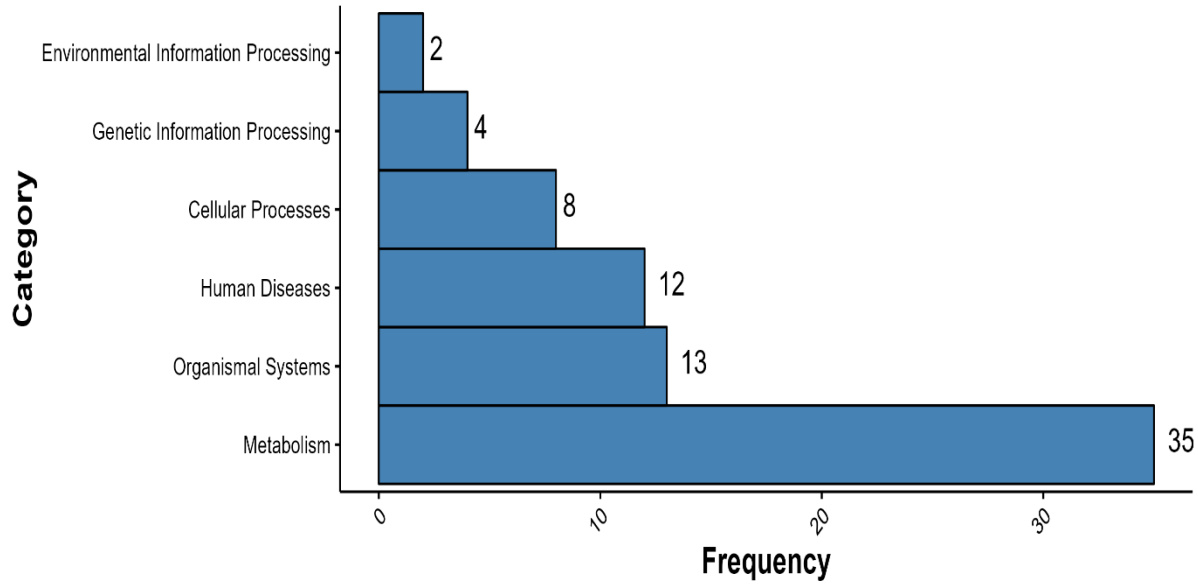


Figure 6.10: **All significantly enriched KEGG pathways detected following the adaption of CHO K1 and DXB11 cell lines from attached to suspension growth. The enriched KEGG pathways have been grouped by KEGG category. Frequency refers to the number of enriched KEGG pathways detected within each category.**

As a theme, metabolism (35/74 pathways) dominates the list of enriched KEGG pathways identified following the adaptation of K1 and DXB11 cells from attached to suspension growth. Cellular processes identified as being enriched included focal adhesion, tight junctions, apoptosis and processes involved in catabolism such as lysosome, phagosome and endocytosis.

As a category metabolism contained the largest number of enriched pathways and was therefore further broken down into various subgroups. Metabolism of carbohydrates was the joint largest single category (10/35 pathways). However, the metabolism of alternative energy sources (lipids and amino acids) dominated (14/35 pathways).

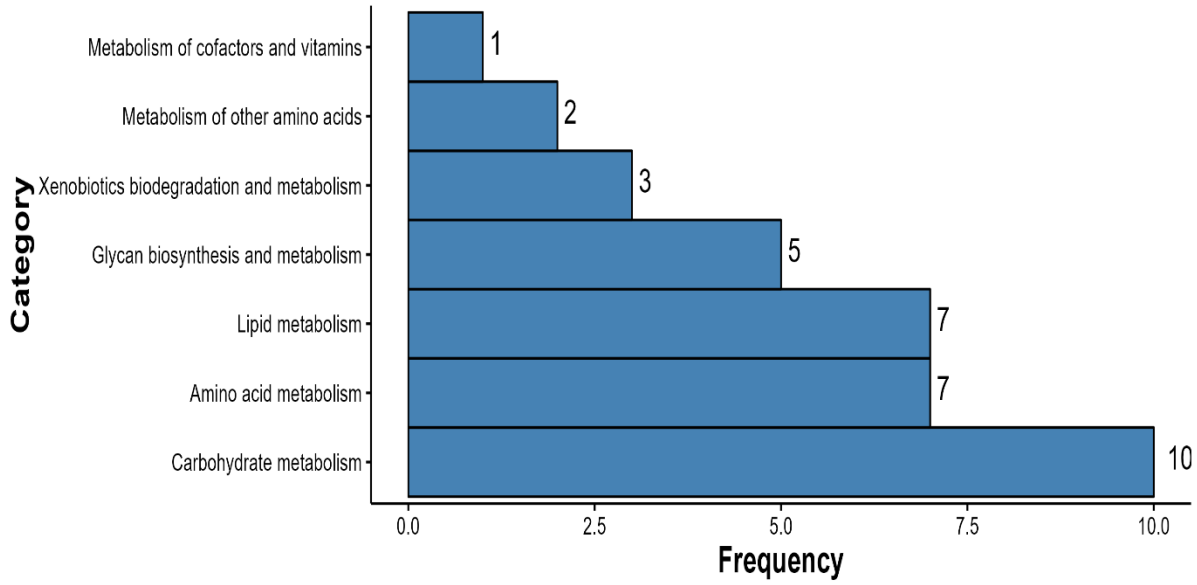


Figure 6.11: **All significantly enriched KEGG pathways detected following the adaption of CHO K1 and DXB11 cell lines from attached to suspension growth which belong to the metabolism category from Figure 6.10. Frequency refers to the number of enriched KEGG pathways detected within each category.**

The ten most enriched pathways ordered by significance (p value) and fold enrichment based on the results of the PSEA analysis of the one hundred common DE proteins identified in this study are displayed in Figure 6.12. Figure 6.12 combines information about the number of proteins detected in each pathway (bubble size) with the degree of enrichment (x-axis) and how significant that pathway is (bubble colour). What it doesn't do is capture whether the proteins identified as being part of those pathways were upregulated or downregulated or indeed what those proteins are. That information is presented in Table 6.4. An additional criterion was applied in the generation of this table. PSEA analysis was performed on the DE proteins from K1 and DXB11 cells separately. This was because although the proteins were common to both cell lines the p value associated with them differed as the two cell lines were analysed in separate MS experiments. Therefore, only KEGG pathways identified as being enriched in both cell lines were reported in this table.

Pathways related to metabolism dominate Figure 6.12 and Table 6.4. As a theme lipid metabolism accounts for 60% of the pathways in Table 6.4 when ranked by significance and 40% when ranked by fold enrichment and the three pathways common to both the ten most significant KEGG pathways and the ten KEGG pathways with the greatest

degree of enrichment. The three pathways common to both the ten most significant KEGG pathways and the 10 KEGG pathways with the greatest degree of enrichment were the *biosynthesis of unsaturated fatty acids, fatty acid degradation* and *bile biosynthesis*. Additionally, glutathione metabolism, propanoate metabolism and insulin signalling were the other pathways listed in Table 6.4 which played a role in lipid metabolism/homeostasis.

One of the attributes of the pathfindR package is that it calculates a significance value for the enriched pathway as a whole. Enriched pathway analysis, in general, compares how many proteins from a given KEGG pathway were identified in the test data set in comparison to a reference dataset. Statistical tests such as a Fishers exact test in conjunction with FDR correction are used to assign a significance (p value) to the pathway as a whole rather than the individual DE proteins. However, the pathfindR packages goes further. It introduces the concept of protein interaction networks (PIN). By mapping individual DE proteins and their associated p value onto a PIN, active subnetworks can be identified. These active subnetworks are then assessed for the presence of enriched KEGG pathways using the PIN as the reference data set. As stated by Ulgen et al. (2019) "Using the genes in the PIN instead of the whole genome provided more statistical strength because active subnetworks are identified using only the genes in the PIN." The default PIN used by the pathfindR package was "Biogrid" and these PINs are updated annually by the package authors (Ulgen et al., 2019).

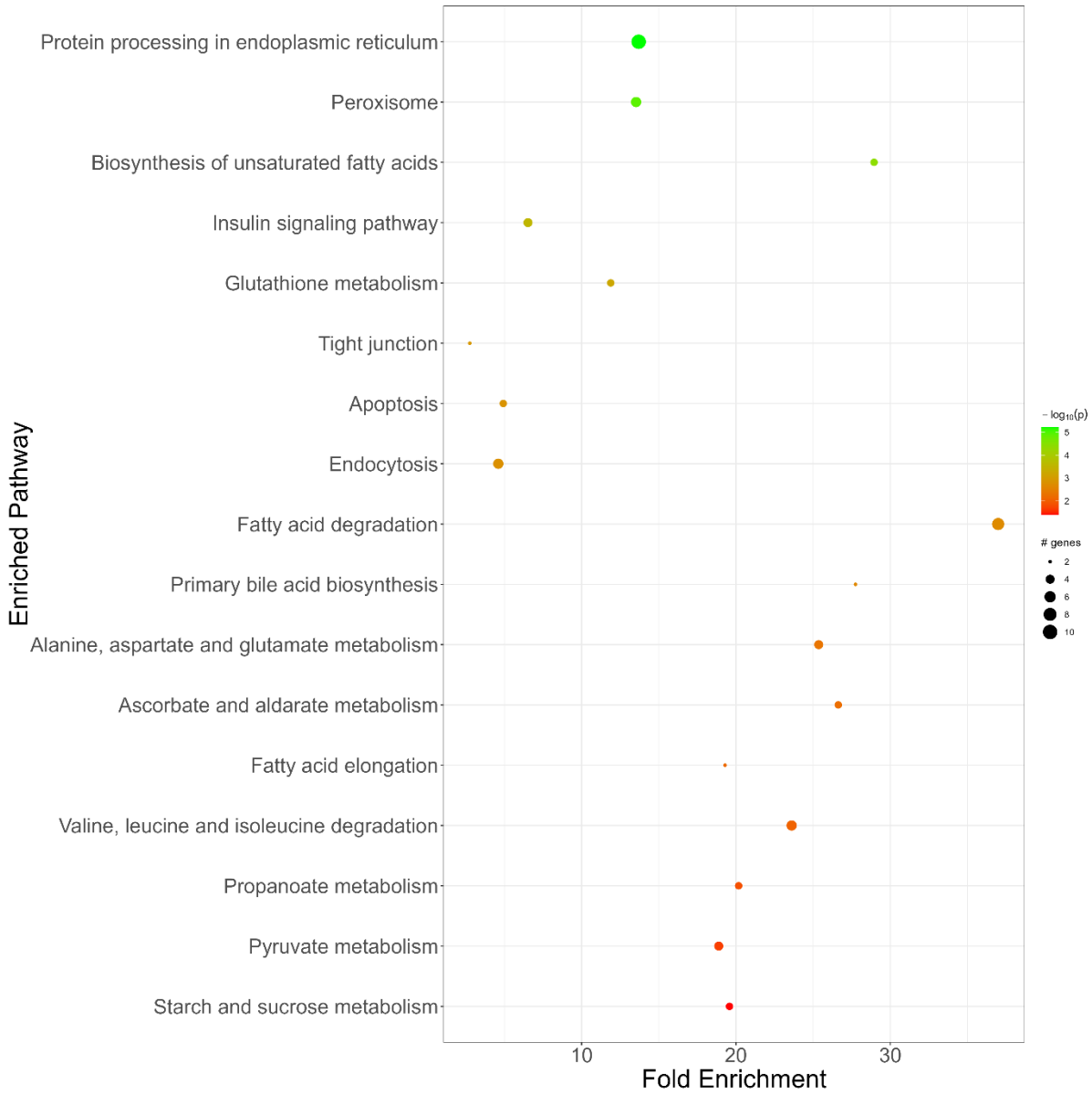


Figure 6.12: **The 10 most significant KEGG pathways and the 10 KEGG pathways with the largest fold enrichment as determined by pathfindR analysis. Bubble size represents the number of differentially expressed proteins detected within each enriched pathway while colour represents statistical significance.**

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated	Down Regulated	Present In
Top 10 KEGG paths ordered by significance						
hsa04141	Protein processing in endoplasmic reticulum	13.7	$5.98E10^{-6}$	RRBP1, HSP90B1, GANAB, LMAN2, SEC24C, PDIA6, TXNDC5	NPLOC4, HSP90AA1, CAPN2	Both
hsa04146	Peroxisome	13.5	$1.21E10^{-5}$	ACOX1, HSD17B4, SCP2, IDH2, PRDX1	NA	Both
hsa01040	Biosynthesis of unsaturated fatty acids	28.9	$3.85E10^{-5}$	ACOX1, HSD17B4, SCP2	NA	Both
hsa04910	Insulin signalling pathway	6.5	$2.97E10^{-4}$	GYS1, PYGL, FLOT1, FASN	NA	Both
hsa00480	Glutathione metabolism	11.9	$5.98E10^{-4}$	GSTP1, IDH2	LAP3	Both
hsa04530	Tight junction	2.8	$1.23E10^{-3}$	NA	TJP1, NEDD4	Both
has04210	Apoptosis	4.9	$1.44E10^{-3}$	CTSD	PARP1, CAPN2	Both
hsa04144	Endocytosis	4.6	$1.66E10^{-3}$	DNM2, CAPZB, SNX1	NEDD4, KIF5B	Both
hsa00071	Fatty acid degradation	37.0	$2.06E10^{-3}$	ACAT1, ACAA2, HADHB, ACOX1, ACADVL, ADH5, ALDH2	NA	Both
hsa00120	Primary bile acid biosynthesis	27.7	$2.37E10^{-3}$	HSD17B4, SCP2	NA	Both
Top 10 KEGG paths ordered by fold enrichment						
hsa00071	Fatty acid degradation	37.0	$4.16E10^{-5}$	ACAT1, ACAA2, HADHB, ACOX1, ACADVL, ADH5, ALDH2	NA	Both

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated	Down Regulated	Present In
hsa01040	Biosynthesis of unsaturated fatty acids	28.9	$6.43E10^{-6}$	ACOX1, HSD17B4, SCP2	NA	Both
hsa00120	Primary bile acid biosynthesis	27.7	$2.37E10^{-3}$	HSD17B4, SCP2	NA	Both
hsa00053	Ascorbate and aldarate metabolism	26.6	$5.92E10^{-3}$	GUSB, AKR1A1, ALDH2	NA	Both
hsa00250	Alanine, aspartate and glutamate metabolism	25.4	$4.70E10^{-3}$	GOT2, ALDH5A1, GLUD1	CAD	Both
hsa00280	Valine, leucine and isoleucine degradation	23.6	$5.88E10^{-5}$	ACAA2, HADHB, PCCB, ALDH2, ACAT1	NA	Both
hsa00640	Propanoate metabolism	20.2	$6.95E10^{-3}$	ACOX1, PCCB, ACAT1	NA	Both
hsa00500	Starch and sucrose metabolism	19.6	$7.38E10^{-3}$	UGP2, GYS1, PYGL	NA	Both
hsa00062	Fatty acid elongation	19.8	$5.00E10^{-3}$	ACAA2, HADHB	NA	Both
hsa00620	Pyruvate metabolism	18.9	$2.13E10^{-3}$	ADH5, AKR1A1, ALDH2, ACAT1	NA	Both
Protein names have not been used for space reasons. Instead, the proteins names associated with all HGNC symbols mentioned in this chapter have been listed in Table 6.12 at the end of this chapter.						

Table 6.4: The ten KEGG pathways displaying the greatest degree of enrichment or significance following PSEA analysis of CHO K1 and CHO DXB11 cells adapted from attached growth to suspension growth in the presence of serum.

6.3.4. Gene Ontology analysis of DE proteins common to CHO K1 and DXB11 cells

Gene ontology (GO) describes the biological processes (BP), molecular functions (MF) and cellular components (CC) associated with an individual gene. Given the link between genes and proteins it can also be used to analyse protein lists. It's designed to be species agnostic and therefore a *Homo sapiens* reference was used as the background dataset for this analysis. The R packages BiomaRt (Durinck et al., 2009) and Annotation Hub (Morgan and Shepard, 2021) were used to assigned gene ontologies to the sixty proteins identified by PSEA and identify which of GO terms were statistically enriched (Appendix 13). Table 6.5 lists the five most significantly enriched GO terms in each ontology associated with the enriched KEGG pathways identified by the PSEA analysis (Table 6.4).

As can been seen from Table 6.5, biological processes dominate the list in terms of significance and suggest that lipid, fatty acid, and organic acid metabolism played a significant role or were particularly perturbed during the adaption of CHO cells from attached to suspension growth. The most significantly enriched cellular locations and molecular functions supported this interpretation. Melanosomes are intracellular organelles which are specifically generated by pigment cells in the skin and eye. It's therefore unlikely they would be present in CHO cells and therefore they were not considered further.

ID	Description	p value	p adjusted	q value	Number of Proteins	Ontology
GO:0042470	Melanosome	6.07E-13	6.16E-11	4.12E-11	10	CC
GO:0048770	Pigment granule	6.07E-13	6.16E-11	4.12E-11	10	CC
GO:0044282	Small molecule catabolic process	1.87E-13	3.62E-10	2.66E-10	16	BP
GO:0046395	Carboxylic acid catabolic process	1.04E-11	1.01E-08	7.43E-09	12	BP
GO:0016054	Organic acid catabolic process	2.10E-11	1.36E-08	9.97E-09	12	BP
GO:0044242	Cellular lipid catabolic process	1.17E-10	5.65E-08	4.15E-08	11	BP
GO:0006631	Fatty acid metabolic process	1.89E-10	7.30E-08	5.37E-08	13	BP
GO:0034774	Secretory granule lumen	3.36E-08	1.62E-06	1.09E-06	10	CC
GO:0060205	Cytoplasmic vesicle lumen	3.77E-08	1.62E-06	1.09E-06	10	CC
GO:0031983	Vesicle lumen	4.00E-08	1.62E-06	1.09E-06	10	CC
GO:0016616	Oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	1.11E-07	1.76E-05	1.24E-05	7	MF
GO:0031406	Carboxylic acid binding	1.26E-07	1.76E-05	1.24E-05	8	MF
GO:0016614	Oxidoreductase activity, acting on CH-OH group of donors	1.92E-07	1.78E-05	1.26E-05	7	MF
GO:0016408	C-acyltransferase activity	4.51E-07	3.13E-05	2.22E-05	4	MF
GO:0051287	NAD binding	7.39E-07	4.11E-05	2.91E-05	5	MF

Table 6.5: **Significantly enriched GO terms following PSEA analysis of CHO K1 and CHO DXB11 cells adapted from attached growth to suspension growth in the presence of serum. Adjusted p value and q Value are the p value following application of the Benjamini-Hochberg formula and false discovery rate (FDR) respectively to the calculated p value.**

6.3.5. Identification of DE proteins and KEGG pathways common to CHO K1 and DXB11 cells which are related to cellular adhesion

The aim behind performing a proteomic analysis of CHO K1 and CHO DXB11 cells following their adaptation from attached growth to suspension growth was to try and

gain an insight into what changes in protein expression correlated with this loss of cellular adhesion. Two models which potentially describe the events and cellular processes involved in the adaption of cells from attached growth to suspension growth are:

- The EMT phenomenon in tumour cells which has been associated with metastasis.
- The reverse process, wound healing whereby soluble factors are recruited to an injury site, where they then clot together and adhere to the surrounding cells.

Recently two public databases of genes associated with these models have been published. These were the EMTome (Vasaikar et al., 2021) and TiRe (Yanai et al., 2016) databases respectively. These databases were downloaded and used to build two R vectors containing three thousand six hundred and thirty-two unique entries from eighty-four publications (EMTome) (Appendix 14) and three hundred and thirty-one unique entries from 311 publications (TiRe) (Appendix 15) respectively. These vectors were then used to filter the list of one hundred common DE genes. The results of this filtering step are presented in Table 6.6 and represent previously published genes associated with EMT and wound healing which were among the 100 DE proteins identified as common to both CHO K1 and CHO DXB11 cells adapted from attached to suspension growth in the presence of 5% FBS.

HGNC Symbol	Protein Name	Uniprot.ID	Status	FC K1	FC DXB11
ACOX1	Acyl-CoA oxidase 1	Q15067	Up	2.376	2.132
AIMP1	Aminoacyl tRNA synthetase complex interacting multifunctional protein 1	Q12904	Down	0.56	0.277
ALDH5A1	Aldehyde dehydrogenase 5 family member A1	P51649	Up	3.068	2.667
ANXA2	Annexin A2	P07355	Up	1.766	1.655
CAPN2	Calpain 2	P17655	Down	0.661	0.183
COLGALT1	Collagen beta(1-O-galactosyltransferase 1	Q8NBJ5	Up	1.833	1.512
COPA	COPI coat complex subunit alpha	P53621	Up	1.507	2.7
CSDE1	Cold shock domain containing E1	O75534	Down	0.519	0.597
CTSD	Cathepsin D	P07339	Up	2.283	2.184
DDX39A	DExD-box helicase 39A	O00148	Down	0.652	0.641
DPYSL2	Dihydropyrimidinase like 2	Q16555	Up	1.979	1.512
FHL2	Four and a half LIM domains 2	Q14192	Down	0.469	0.529
FLNA	Filamin A	P21333	Down	0.655	0.103
FLOT1	Flotillin 1	O75955	Up	2.284	1.647
GANAB	Glucosidase II alpha subunit	Q14697	Up	1.673	1.639
GPD1L	Glycerol-3-phosphate dehydrogenase 1 like	Q8N335	Up	2.883	2.754
HSP90AA1	Heat shock protein 90 alpha family class A member 1	P07900	Down	0.606	0.148
IDH2	Isocitrate dehydrogenase (NADP(+)) 2	P48735	Up	1.993	1.714
IMPDH2	Inosine monophosphate dehydrogenase 2	P12268	Down	0.523	0.387

HGNC Symbol	Protein Name	Uniprot.ID	Status	FC K1	FC DXB11
LAMA5	Laminin subunit alpha 5	O15230	Down	0.459	0.191
LRRFIP1	LRR binding FLII interacting protein 1	Q32MZ4	Down	0.187	0.3
LUC7L2	LUC7 like 2, pre-mRNA splicing factor	Q9Y383	Down	0.487	0.467
NCAM1	Neural cell adhesion molecule 1	P13591	Down	0.375	0.332
NNT	Nicotinamide nucleotide transhydrogenase	Q13423	Up	2.18	1.807
P3H1	Prolyl 3-hydroxylase 1	Q32P28	Up	2.365	16.219
PARP1	poly(ADP-ribose) polymerase 1	P09874	Down	0.536	0.635
PDIA6	Protein disulfide isomerase family A member 6	Q15084	Up	2.156	1.713
PLA2G4A	Phospholipase A2 group IVA	P47712	Down	0.443	0.346
PLIN3	Perilipin 3	O60664	Up	1.964	1.836
PPIB	Peptidylprolyl isomerase B	P23284	Up	1.842	1.633
PRDX1	Peroxiredoxin 1	Q06830	Up	1.544	2.204
PRDX2	Peroxiredoxin 2	P32119	Up	2.608	1.759
PSME1	Proteasome activator subunit 1	Q06323	Up	2.011	1.797
RRBP1	Ribosome binding protein 1	Q9P2E9	Up	1.873	1.616
TJP1	Tight junction protein 1	Q07157	Down	0.547	0.469

Table 6.6: DE proteins identified as being common to both K1 and DXB11 cells after adaption from attached to suspension growth in the presence of serum which are also known to be involved in EMT.

Thirty-five proteins identified as being common to both K1 and DXB11 cells in this study had been previously published by other groups as being associated with the EMT phenotype. This represented over a third of the DE proteins identified in this study as being common to K1 and DXB11 cells after adaptation from attached to suspension growth in the presence of 5% serum and 58% of the 60 DE proteins identified as being present in the enriched KEGG paths.

HGNC Symbol	Protein Name	Uniprot.ID	Status	FC K1	FC DXB11
LGALS3BP	Galectin 3 binding protein	Q08380	Up	3.725	4.980
ARHGAP1	Rho GTPase activating protein 1	Q07960	Up	1.624	1.652
FHL2	Four and a half LIM domains 2	Q14192	Down	0.469	0.529
AIMP1	Aminoacyl tRNA synthetase complex interacting multifunctional protein 1	Q12904	Down	0.560	0.277
HSP90AA1	Heat shock protein 90 alpha family class A member 1	P07900	Down	0.606	0.148
FLNA	Filamin A	P21333	Down	0.655	0.103

Table 6.7: **DE proteins identified as being common to both K1 and DXB11 cells after adaption from attached to suspension growth which are also known to be involved in wound healing.**

In comparison to the EMT data (Table 6.6) only six proteins identified in this study as being common to both K1 and DXB11 cells following adaptation from attached to suspension growth had been previously published by other groups as have an association with the biological process of wound healing (Table 6.7).

Four proteins were present in both Tables 6.6 and 6.7. These were *four and a half LIM domains 2 (FHL2), aminoacyl tRNA synthetase complex interacting multifunctional protein (AIMP1), heat shock protein 90 alpha family class A member 1 (HSP90AA1)* and *filamin A (FLNA)*. One of these, filamin A, is directly involved in attachment via its role in the formation of focal adhesions. Two others aminoacyl tRNA synthetase complex interacting multifunctional protein 1, heat shock protein 90 alpha family class A member 1 are known to be associated with the process of cellular adhesion.

To gain further insight into the proteins/KEGG pathways involved in cellular adhesion, the entire KEGG database was searched. The results of this search were used to build an R vector containing forty-nine individual entries (Appendix 16) which was then used to filter the list of sixty-six statistically enriched KEGG pathways listed in Appendix 12. The results of this filtering step are presented in Table 6.8 and represent the adhesion related KEGG pathways which were statistically enriched during the adaption of CHO K1 and CHO DXB11 cells from attached to suspension growth.

Six unique KEGG pathways (Table 6.8) associated with adhesion were found to be statistically enriched during adaption of CHO K1 and DXB11 cells from attached to suspension growth. The pathway with the greatest fold enrichment was the estrogen signalling pathway while the most significantly enriched pathway was tight junctions. Focal adhesions were the other type of cell attachment which was enriched. Ten unique proteins from these six pathways were detected in this analysis. Two of these were found to be upregulated while the other eight were downregulated. All proteins detected in the tight junction and focal adhesion pathways were down regulated. Both upregulated proteins were components of the Estrogen pathway. Proteins encoded by the genes *PARP1, TJP1, PLA2G4A, LAMA5, FLNA, CAPN2, HSP90AA1* were all previously identified as proteins involved in EMT.

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated DE Proteins	Down Regulated DE Proteins
hsa04915	Estrogen signalling pathway	4.86	$8.56\ E10^{-3}$	HSP90B1, CTSD	HSP90AA1
hsa04370	VEGF signalling pathway	3.76	$2.25E10^{-2}$	None	PLA2G4A
hsa04510	Focal adhesion	3.36	$1.26E10^{-2}$	None	LAMA5, CAPN2, FLNA
hsa04530	Tight junction	2.76	$1.22E10^{-2}$	None	TJP1, NEDD4
hsa04064	NF-kappa B signalling pathway	2.13	$4.21E10^{-2}$	None	PARP1
hsa05417	Lipid and atherosclerosis	2.07	$1.58E10^{-2}$	HSP90B1	HSP90AA1
Protein names have not been used for space reasons. Instead, the proteins names associated with all HGNC symbols mentioned in this chapter have been listed in Table 6.12 at the end of this chapter.					

Table 6.8: **All adhesion related KEGG pathways found to be significantly enriched following adaption of CHO K1 and CHO DXB11 cells from attached to suspension growth in the presence of 5% FBS.**

6.3.6. Gene Ontology analysis of adhesion related DE proteins common to CHO K1 and DXB11 cells

GO analysis revealed seventy-seven GO terms to be significantly enriched (Appendix 13). Table 6.9 lists the ten most significantly enriched GO terms for each ontology associated with the ten proteins listed in Table 6.8.

ID	Description	p value	p adjusted	q value	No of Proteins associated with the GO Term	Ontology
GO:0042470	Melanosome	$1.82E10^{-5}$	$7.54E10^{-4}$	$3.35E10^{-4}$	3	CC
GO:0048770	Pigment granule	$1.82E10^{-5}$	$7.54E10^{-4}$	$3.35E10^{-4}$	3	CC
GO:0071682	Endocytic vesicle lumen	$4.47E10^{-5}$	$1.23E10^{-4}$	$5.48E10^{-4}$	2	CC
GO:0044295	Axonal growth cone	$8.87E10^{-5}$	$1.84E10^{-3}$	$8.17E10^{-4}$	2	CC
GO:0005938	Cell cortex	$4.02E10^{-4}$	$6.67E10^{-3}$	$2.96E10^{-3}$	3	CC
GO:0005925	Focal adhesion	$1.03E10^{-3}$	$9.98E10^{-3}$	$4.43E10^{-3}$	3	CC
GO:0043202	Lysosomal lumen	$1.05E10^{-3}$	$9.98E10^{-3}$	$4.43E10^{-3}$	2	CC
GO:0030055	Cell-substrate junction	$1.08E10^{-3}$	$9.98E10^{-3}$	$4.43E10^{-3}$	3	CC
GO:0062023	Collagen-containing extracellular matrix	$1.08E10^{-3}$	$9.98E10^{-3}$	$4.43E10^{-3}$	3	CC
GO:0030863	Cortical cytoskeleton	$1.23E10^{-3}$	$1.02E10^{-2}$	$4.53E10^{-3}$	2	CC
GO:0150105	Protein localization to cell-cell junction	$5.28E10^{-5}$	$1.96E10^{-2}$	$9.23E10^{-3}$	2	BP[1]
GO:0051017	Actin filament bundle assembly	$6.14E10^{-5}$	$1.96E10^{-2}$	$9.23E10^{-3}$	3	BP[1]
GO:0061572	Actin filament bundle organization	$6.63E10^{-3}$	$1.96E10^{-2}$	$9.23E10^{-3}$	3	BP[1]
GO:0046332	SMAD binding	$7.66E10^{-4}$	$4.11E10^{-2}$	$1.75E10^{-2}$	2	MF
GO:0042826	Histone deacetylase binding	$1.64E10^{-3}$	$4.11E10^{-2}$	$1.75E10^{-2}$	2	MF

ID	Description	p value	p adjusted	q value	No of Proteins associated with the GO Term	Ontology
GO:0099106	Ion channel regulator activity	$1.76E10^{-3}$	$4.11E10^{-2}$	$1.75E10^{-2}$	2	MF
GO:0051082	Unfolded protein binding	$1.85E10^{-3}$	$4.11E10^{-2}$	$1.75E10^{-2}$	2	MF
GO:0016247	Channel regulator activity	$2.68E10^{-3}$	$4.50E10^{-2}$	$1.92E10^{-2}$	2	MF
GO:0102545	Phosphatidyl phospholipase B activity	$5.98E10^{-3}$	$4.50E10^{-2}$	$1.92E10^{-2}$	1	MF
GO:0051020	GTPase binding	$4.50E10^{-3}$	$4.50E10^{-2}$	$1.92E10^{-2}$	2	MF
GO:0004198	Calcium-dependent cysteine-type endopeptidase activity	$8.69E10^{-3}$	$4.50E10^{-2}$	$1.92E10^{-2}$	1	MF
GO:0047498	Calcium-dependent phospholipase A2 activity	$8.69E10^{-3}$	$4.50E10^{-2}$	$1.92E10^{-2}$	1	MF
GO:1990404	Protein ADP-ribosylase activity	$8.69E10^{-3}$	$4.50E10^{-2}$	$1.92E10^{-2}$	1	MF

[1] There were only three biological processes listed among the seventy-seven adhesion associated enriched GO terms

Table 6.9: List of significantly enriched GO terms associated with the DE proteins listed in Table 6.8.

It's important to note that although the seventy-seven GO terms listed in Appendix 13 were derived from DE expressed proteins identified as playing a role in pathways relating to adhesion these GO terms will cover the full range of biological events, not just adhesion. This is because, a single protein can be associated with multiple GO terms. To create a list of GO terms related to adhesion the entire GO environment was searched for terms related to cellular adhesion. These terms were used to build an R vector which contained nine hundred and thirty-two unique entries (Appendix 17) which was used to further refine the seventy-seven enriched GO terms listed in Appendix 13. This filtering step identified eleven significantly enriched adhesion specific GO terms associated with the ten proteins identified as being part of the adhesion related enriched KEGG pathways listed in Table 6.8. These eleven enriched GO terms are presented in Table 6.10.

The output of this filtering step suggested that disruption of focal adhesions, cadherin binding and membrane rafts played a role in the adaption of CHO K1 and DXB11 cells from attached to suspension growth. It's known that focal adhesion and integrins participate in the anchoring of cells to the ECM and that the loss of integrin-mediated adhesion has been associated with caveolin-1 dependent internalisation and recycling of membrane rafts. Likewise, adherens junctions and cadherins participate in intercellular contacts between neighbouring cells. Therefore, the disruption of these structures would appear to be a key step in the adaption of cells from attached growth to suspension growth.

ID	Description	p value	p adjusted	q value	No. of Proteins	Ontology
GO:0005938	Cell cortex	$4.02E10^{-4}$	$6.67E10^{-3}$	$2.96E10^{-3}$	3	CC
GO:0005925	Focal adhesion	$1.03E10^{-3}$	$9.98E10^{-3}$	$4.43E10^{-2}$	3	CC
GO:0062023	Collagen-containing extracellular matrix	$1.08E10^{-3}$	$9.98E10^{-3}$	$4.43E10^{-2}$	3	CC
GO:0150105	Protein localization to cell-cell junction	$5.28E10^{-5}$	$1.96E10^{-2}$	$9.23E10^{-2}$	2	BP
GO:0030139	Endocytic vesicle	$1.02E10^{-2}$	$3.21E10^{-2}$	$1.42E10^{-2}$	2	CC
GO:0016327	Apicolateral plasma membrane	$1.07E10^{-2}$	$3.21E10^{-2}$	$1.42E10^{-2}$	1	CC
GO:0045121	Membrane raft	$1.13E10^{-2}$	$3.21E10^{-2}$	$1.42E10^{-2}$	2	CC
GO:0043198	Dendritic shaft	$1.63E10^{-2}$	$4.09E10^{-2}$	$1.82E10^{-2}$	1	CC
GO:0046332	SMAD binding	$7.66E10^{-4}$	$4.11E10^{-2}$	$1.75E10^{-2}$	2	MF
GO:0045296	Cadherin binding	$1.34E10^{-2}$	$4.50E10^{-2}$	$1.92E10^{-2}$	2	MF
GO:0016248	Channel inhibitor activity	$2.11E10^{-2}$	$5.17E10^{-2}$	$2.20E10^{-2}$	1	MF

Table 6.10: **All adhesion related GO terms found to be significantly enriched following adaption of CHO K1 and CHO DXB11 cells from attached to suspension growth in the presence of 5% FBS. Adjusted p value and q Value are the p value following application of the Benjamini-Hochberg formula and false discovery rate (FDR) respectively to the calculated p value.**

6.3.7. Impact of correcting for false discovery rates (FDR) on the identification of differentially expressed proteins common to CHO K1 and CHO DXB11 cells after adaptation to suspension growth in media containing 5% FBS.

During the review of this thesis prior to submission the lack of an FDR filter during selection of DE proteins in the Progenesis QI for Proteomics workflow was commented upon by the project sponsor. While Progenesis QI for Proteomics version 2.0 can apply an FDR criterion at both the DE peptide and DE protein selection stage, one was not applied in the original analysis. Given that the application of an FDR criteria to a dataset increases the proportion of false negatives associated with that dataset, it may not have been appropriate to apply an FDR filter to both the DE peptides and protein lists as well as the subsequent bioinformatics analysis. Therefore, an FDR value of 0.01 was used during protein assignment using the Proteome Discoverer package and an FDR filter in the form of a q value cut-off of ≤ 0.05, as determined using the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995), was applied during the identification of enriched gene ontology terms.

The omission of an FDR criteria at the peptide level meant that the number of peptides which were identified as DE were maximised. FDR criteria were applied to all other statistical analysis in order to provide additional confidence in the proteins/KEGG pathways which were reported as enriched. However, the DE protein list generated by Progenesis QI for Proteomics has been interrogated following application of a retrospective FDR criteria at this step and an assessment of the impact omitting such a criteria had in the identification of DE proteins has been made. This assessment is discussed in the following paragraphs.

One thousand six hundred and seventy-eight DE proteins were detected in CHO DXB11 cells compared to one thousand two hundred and eighty-five in CHO K1 cells following adaptation to suspension growth in serum free media.

Following application of a false discovery rate (q value) of 0.01, the number of DE proteins identified in CHO K1 cells was reduced to four hundred and thirty-four (Appendix 18) and to one hundred and seventeen in DXB11 cells (Appendix 19). An Upset plot identified twenty DE proteins common to both CHO K1 and CHO DXB11 cells.

These twenty DE proteins were analysed pathways by peptide set enrichment analysis (PSEA) in order to identify enriched KEGG pathways they were implicated in. Enriched gene ontology terms associated with these twenty DE proteins were identified using the Annotation Hub package in R.

Table 6.11 lists the top five proteins in either cell line based on:

a. Absolute fold change (absolute FC) between the two experimental groups
b. p value

However, it was decided to focus on the twenty three DE proteins, identified using an Upset plot (Figure 6.13), which are either upregulated (Up.DXB11 and Up.K1) or downregulated (Down.DXB11 and Down.K1) in both CHO K1 and CHO DXB11 cells. These twenty three proteins are more applicable to explaining the events underpinning the adaptation of CHO K1 and CHO DXB11 cells from attached to suspension growth in the presence of serum.

Cell Line	Protein Name	HGNC Symbol	p Value	Max FC[1]	Log_2(Absolute FC)[2]	Present in all attached reps
K1	Prothymosin alpha 1	PTMA	$3.62E10^{-2}$	72.4	-6.18	Yes
	40S ribosomal protein S3a	RPS3A	$2.40E10^{-2}$	8.22	-3.04	Yes
	Mitochondrial ferrochelatase	FECH	$2.96E10^{-3}$	7.23	2.85	Yes
	3 hydroxybutyrate dehydrogenase type 2	BDH2	$9.36E10^{-4}$	6.67	2.73	Yes
	Integrin alpha 7 precursor	ITGA7	$2.57E10^{-3}$	5.62	2.49	Yes
DXB11	3'-phosphoadenosine 5'-phosphosulfate synthase 1	PAPSS1	$1.33E10^{-2}$	21.9	4.45	Yes
	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1 like	MTHFD1L	$1.85E10^{-3}$	13.1	3.71	Yes
	UDP-N-acetylhexosamine pyrophosphorylase- like protein 1	UAP1L1	$7.25E10^{-4}$	7.50	2.91	Yes
DXB11	Calpain-2 catalytic subunit precursor	CAPN2	$4.12E10^{-3}$	5.45	-2.45	Yes
	Laminin subunit alpha 5 precursor	LAMA5	$3.40E10^{-3}$	5.23	-2.39	Yes
K1	Exportin-5	XP05	$8.74E10^{-4}$	1.78	0.83	Yes

	Von Willebrand factor A domain containing protein 8	VWA8	$8.74E10^{-4}$	1.78	0.83	Yes
	2-oxoglutarate and iron-dependent oxygenase domain-containing protein 3	OGFOD3	$8.77E10^{-4}$	4.24	2.09	Yes
	Lysophospholipid acyltransferase 7	MBOAT7	$8.83E10^{-4}$	3.55	1.83	Yes
	Early growth response protein 1	ERIGC1	$8.85E10^{-4}$	2.11	1.08	Yes
DXB11	Filamin B	FLNB	$1.71E10^{-4}$	1.71	0.77	Yes
	Neuroblast differentiation associated protein	AHNAK	$2.37E10^{-4}$	1.62	0.70	Yes
	Scavenger receptor class B member 1	SCARB1	$4.04E10^{-4}$	5.16	2.37	Yes
	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	SDHA	$4.44E10^{-4}$	2.16	1.11	Yes
DXB11	Coatomer subunit alpha	COPA	$6.92E10^{-4}$	2.70	1.43	Yes

Table 6.11: **The five most significant DE proteins detected in CHO K1 and CHO DXB11 cell lines as ranked by max fold change and statistical significance following application of a q value ≤ 0.01 at the DE protein level in the Progenesis QI for Proteomics workflow.**

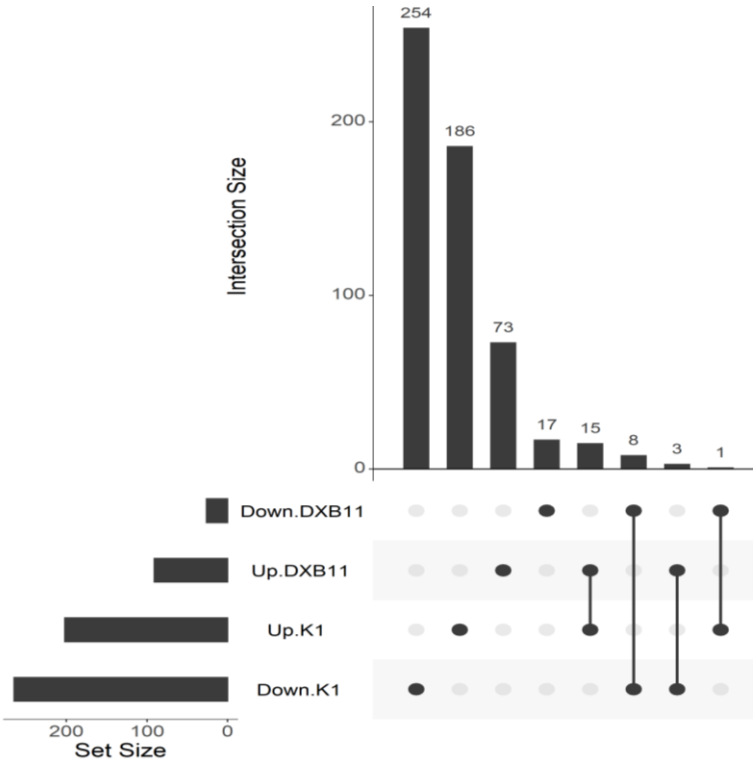


Figure 6.13: **Number of unique DE proteins present in each of the subpopulations of interest and the number of DE proteins which are common to two or more subpopulations. The numbers above the bars refer to the number of DE proteins represented by that bar. Set Size on the lower left-hand corner of the figure refers to the total number of DE proteins in each subpopulation (e.g., there are 254 DE proteins in the Down.K1 subpopulation and 17 DE proteins in the Down.DXB11 subpopulation. However, these two subpopulations have 8 DE proteins in common. Subpopulations which have no DE proteins in common eg Up.DXB11 and Down.DXB11 are not represented in Figure 6.13.**

Figures 6.14 and 6.15 show the distribution of the nine statistically enriched KEGG pathways detected following the adaptation of K1 and DXB11 cell lines from attached to suspension growth in media containing serum across the seven themes which comprise the KEGG system and within the metabolic theme respectively.

As a theme, metabolism (8/18 pathways) and organismal systems (6/18 pathways) dominated the list of enriched KEGG pathways identified following the adaptation of K1 and DXB11 cells from attached to suspension growth.

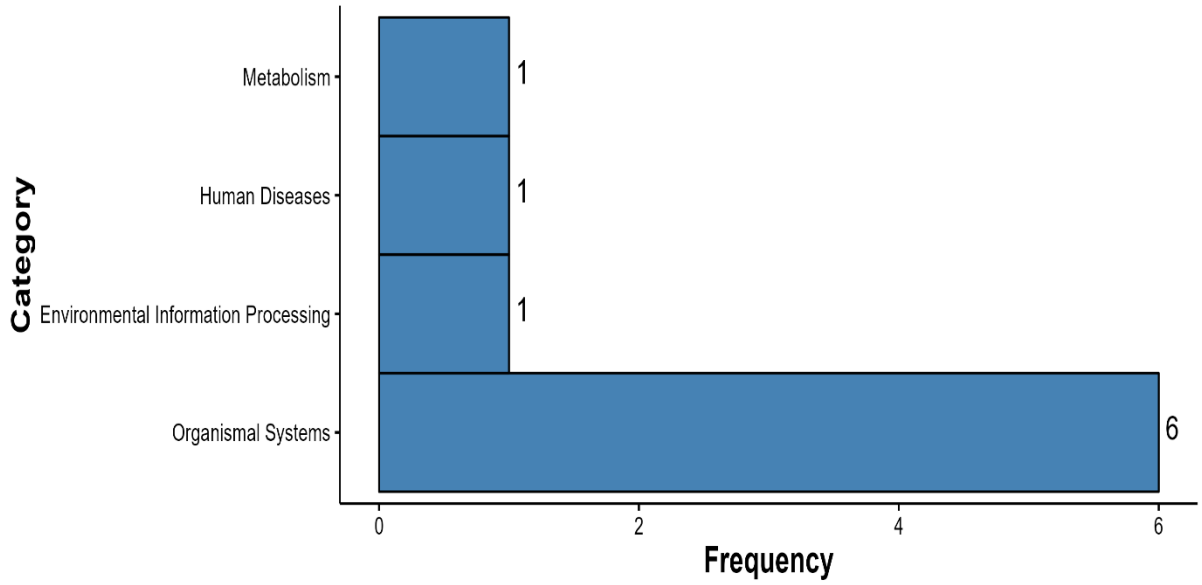


Figure 6.14: **Categorisation of significantly enriched KEGG pathways detected following the adaption of CHO K1 and DXB11 cell lines from suspension growth media containing serum to suspension growth in serum free media. The enriched KEGG pathways have been grouped by KEGG category. Frequency refers to the number of enriched KEGG pathways detected within each category.**

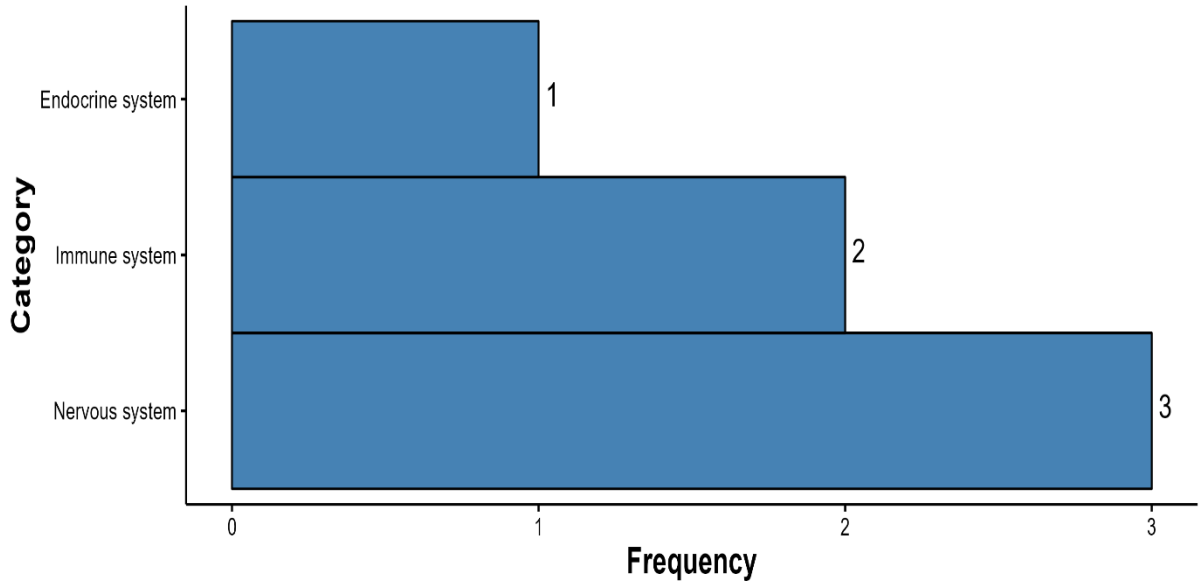


Figure 6.15: **All significantly enriched KEGG pathways detected following the adaption of CHO K1 and DXB11 cell lines from suspension growth in media containing serum to suspension growth in serum free media which belong to the organismal systems category from Figure 6.14. Frequency refers to the number of enriched KEGG pathways detected within each category.**

Figures 6.14 and 6.15 suggest that the changes resulting from the adaption to suspension growth were systemic rather than metabolic. These included enrichment of pathways involved in signalling and phagocytosis.

All enriched KEGG pathways based on the results of the PSEA analysis of the twenty common DE proteins identified in this study, following the application of an FDR cut-off, are displayed in Figure 6.16. Figure 6.16 combines information about the number of proteins detected in each pathway (bubble size) with the degree of enrichment (x-axis) and how significant that pathway is (bubble colour). What it doesn't do is capture whether the proteins identified as being part of those pathways were upregulated or downregulated or indeed what those proteins are. That information is presented in Table 6.12.

PSEA analysis of these twenty DE proteins identified nine KEGG pathways (Figure 6.16) as being enriched. Two DE proteins *cytosolic phospholipase A2 (PLA2G4A)* and *nicotinamide nucleotide transhydrogenase (NNT)* were identified as being a part of these pathways. Filtering of these nine pathways with an R vector containing forty-nine KEGG paths known to be associated with adhesion (Appendix 16) revealed the adhesion related KEGG pathways which were statistically enriched during the adaption of CHO K1 and CHO DXB11 cells from attached to suspension growth (Table 6.13).

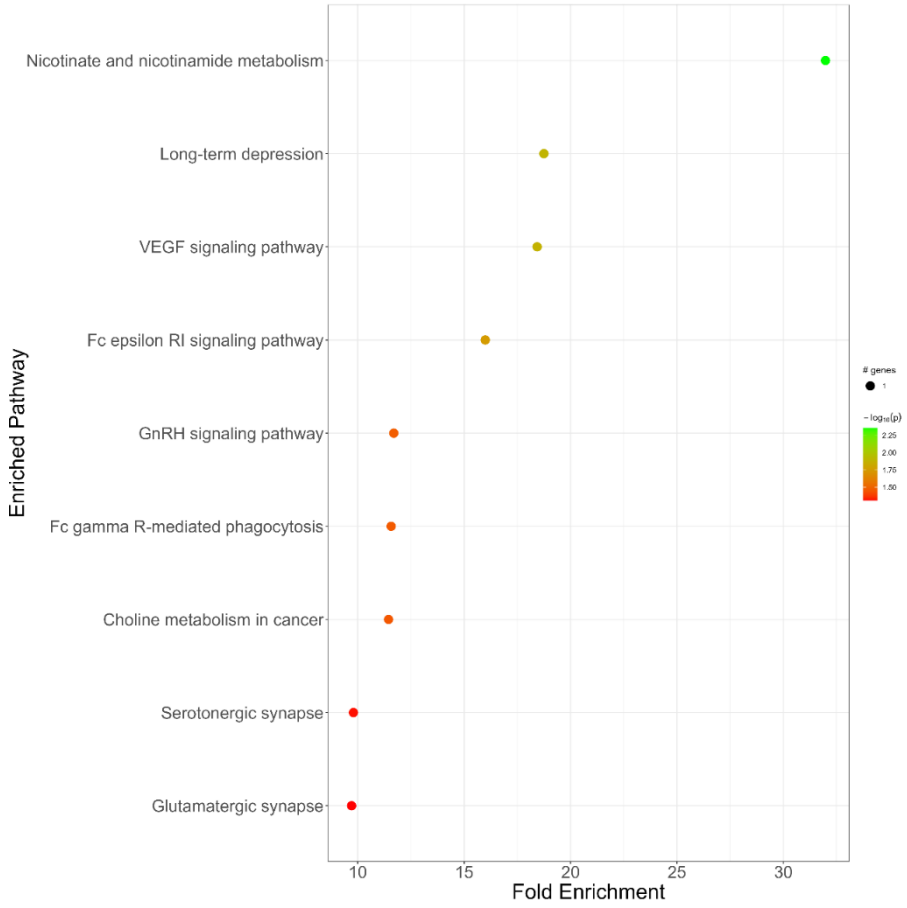


Figure 6.16: **All DE KEGG pathways detected following adaptation of CHO K1 and CHO DXB11 cells to suspension growth. Bubble size represents the number of differentially expressed proteins detected within each enriched pathway while colour represents statistical significance.**

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated	Down Regulated
hsa00760	Nicotinate and nicotinamide metabolism	31.98088	0.004433	NNT	None
hsa04730	Long-term depression	18.74741	0.013042	None	PLA2G4A
hsa04370	VEGF signaling pathway	18.42966	0.013499	None	PLA2G4A
hsa04664	Fc epsilon RI signaling pathway	15.99044	0.017962	None	PLA2G4A
hsa04912	GnRH signaling pathway	11.69194	0.03368	None	PLA2G4A
hsa04666	Fc gamma R-mediated phagocytosis	11.56755	0.03441	None	PLA2G4A
hsa05231	Choline metabolism in cancer	11.44579	0.035148	None	PLA2G4A
hsa04726	Serotonergic synapse	9.795946	0.04801	None	PLA2G4A
hsa04724	Glutamatergic synapse	9.708482	0.04888	None	PLA2G4A

Table 6.12: **All KEGG pathways following PSEA analysis of CHO K1 and CHO DXB11 cells adapted from growth in media containing serum to growth in serum free media. An additional criterion was applied in the generation of this table. PSEA analysis was performed on the DE proteins from K1 and DXB11 cells separately. This was because although the proteins were common to both cell lines the p value associated with them differed as the two cell lines were analysed in separate MS experiments. Therefore, only KEGG pathways identified as being enriched in both cell lines were reported in this table. For space reasons HGNC symbols rather than protein names have been used throughout this table. Protein names are listed in Table 6.12 at the end of this chapter.**

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated	Down Regulated
hsa04370	VEGF signaling pathway	18.4	0.01	None	PLA2G4A

Table 6.13: **All significantly enriched adhesion KEGG pathways common to CHO K1 and CHO DXB11 cells adapted from attached to suspension growth in media containing serum Only KEGG pathways identified as being enriched in both cell lines were reported in this table. For space reasons HGNC symbols rather than protein names have been used. Protein names are listed in Table 6.12 at the end of this chapter.**

Only one KEGG pathway (Table 6.13) associated with adhesion was found to be significantly enriched. However, phagocytosis was previously identified as an enriched pathway (Table 6.12). Phagocytosis, in its simplest terms, is the engulfing of a foreign particle by a cell and is most commonly associated with the innate immune system. However, consideration of the steps involved in phagocytosis such as the binding of a foreign particle to the cell and reorganisation of the cells cytoskeleton to allow engulfing of the foreign particle also occur during cellular adhesion. Indeed, a paper by May and Machesky published in 2001 uses focal adhesions as a model for describing phagocytosis.

Gene ontology analysis of the twenty differentially expressed proteins present in the enriched KEGG pathways revealed an association with two hundred and seventy four different GO terms (Appendix 27). Table 6.14 lists the five most significantly enriched BP and MF GO terms as well as all four CC GO terms associated with the enriched KEGG pathways identified by the PSEA analysis (Table 6.12).

ID	Description	p value	p adjusted	q value	# Proteins	Ontology
GO:0009201	ribonucleoside triphosphate biosynthetic process	1.93E-06	0.0005	0.0002	3	BP
GO:0009142	nucleoside triphosphate biosynthetic process	2.94E-06	0.0005	0.0002	3	BP
GO:0009199	ribonucleoside triphosphate metabolic process	3.27E-06	0.0005	0.0002	3	BP
GO:0009141	nucleoside triphosphate metabolic process	6.46E-06	0.0008	0.0003	3	BP
GO:0009209	pyrimidine ribonucleoside triphosphate biosynthetic process	1.24E-05	0.001	0.0004	2	BP
GO:0097001	ceramide binding	2.13E-05	0.001	0.0003	2	MF
GO:0046625	sphingolipid binding	4.70E-05	0.001	0.0003	2	MF
GO:0016879	ligase activity, forming carbon-nitrogen bonds	0.0001	0.003	0.0005	2	MF
GO:0051287	NAD binding	0.0002	0.003	0.0005	2	MF
GO:0004553	hydrolase activity, hydrolysing O-glycosyl compounds	0.0006	0.007	0.001	2	MF
GO:0043202	lysosomal lumen	0.0005	0.01	0.007	2	CC
GO:0005766	primary lysosome	0.001	0.01	0.007	2	CC
GO:0042582	azurophil granule	0.001	0.01	0.007	2	CC
GO:0005775	vacuolar lumen	0.002	0.01	0.007	2	CC

Table 6.14: **Significantly enriched GO terms following PSEA analysis of CHO K1 and CHO DXB11 cells adapted from growth in media containing serum to growth in serum free media. Adjusted p value and q Value are the p value following application of the Benjamini-Hochberg formula and false discovery rate (FDR) respectively to the calculated p value.**

As can been seen from Table 6.14, GO terms associated with biological processes dominate the list in terms of significance and suggest that the synthesis and metabolism of nucleosides predominates the biological processes underpinning the adaptation of CHO cells from attached to suspension growth in media containing serum. The list of

most significantly enriched cellular components was dominated by regions of the cell which support lytic reactions such as the lysosome or azurophil granule. This observation is supported by the GO terms describing enriched molecular functions, all of which describe the formation or breaking of chemical bonds. Lysis of an existing metabolite or chemical bond is often the precursor to the formation of a new one. Many of these GO terms provide further support the identification of phagocytosis as an enriched KEGG pathway.

6.4. Discussion

This study compared significantly ($p \leq 0.05$ and fold change ≥ 1.5) differentially expressed proteins detected in cells grown attached to a surface or in suspension. Two different cell lines CHO K1 and CHO DXB11 were used. Although the total number of differentially expressed proteins detected differed between CHO K1 and CHO DXB11 cell lines, one hundred differentially expressed proteins were common to both CHO cell lines. A variety of bioinformatic tools were used to identify enriched KEGG pathways and gene ontology terms to identify an overall biological theme describing these one hundred common differentially expressed proteins.

6.4.1. Reproducibility of the number of proteins identified in each replicate and the impact on the identification of enriched KEGG pathways

In excess of three thousand proteins were identified in each replicate for both K1 experimental groups and the DXB11 5% FBS attached experimental group. However, in the DXB11 5% suspension experimental group greater variability was observed between replicates with the number of proteins detected ranging from one thousand nine hundred and sixty seven to four thousand one hundred and fifty seven. Despite this variability only two proteins from the list of common enriched KEGG pathways (Table 6.4) were found to be not present in all sample replicates. These two proteins were (*E3 ubiquitin protein ligase NEDD4 and aldehyde dehydrogenase 5 family member A1*) and three separate KEGG pathways (Tight Junction, Endocytosis and Alanine, aspartate and glutamate metabolism) were impacted. It unlikely that the identification of the endocytosis or alanine, aspartate and glutamine metabolism pathways as enriched

would be changed by excluding E3 ubiquitin protein ligase NEDD4 or aldehyde dehydrogenase 5 family member A1 from the analysis given that several other proteins were detected in these pathways. However, the picture is not so clear cut for the tight junction pathway. Only two proteins were identified in this pathway, E3 ubiquitin protein ligase NEDD4 and tight junction protein 1. However, given that E3 ubiquitin protein ligase NEDD4 was found to be present in all four replicates of the DXB11 5% FBS attached cells and was only present in one replicate of the DXB11 5% FBS suspension cells it was therefore considered that this was a true differentially expressed protein and that the identification of the tight junction pathway as enriched was appropriate.

6.4.2. Adaptation from attached to suspension growth and metabolic pathways

The analysis presented in Figures 6.10 and 6.11 suggests that pathways involved in metabolism and specifically the metabolism of carbohydrates, amino acids and lipids were all significantly enriched during the adaption of CHO K1 and CHO DXB11 cells from attached growth to suspension growth. These macromolecules have many functions within the cell but the function they all have in common is that they can be used as a source of energy. Eukaryotic cells typically obtain their energy from glucose via glycolysis, the products of which feed the TCA cycle and oxidative phosphorylation However, in the absence of glucose, eukaryotic cells can generate energy (ATP) via alternative sources. For example:

- valine, leucine, and isoleucine are all branched amino acids, Leucine and isoleucine can be catabolised into acetyl-CoA while valine can be catabolised into succinyl-CoA. Both acetyl-CoA and succinyl-CoA are key intermediates in the TCA cycle.
- fatty acids are catabolised via β-oxidation to either acetyl-CoA or propionyl-CoA. Propionyl-CoA is subsequently catabolised via the propanoate metabolism pathway to succinyl-CoA.
- the insulin signalling pathway is known to regulate the catabolism of fatty acids under specific conditions.

The examples listed above are all examples of metabolic pathways which were found to be significantly enriched during the adaption of K1 and DXB11 cells from attached to suspension growth. The fact that these KEGG pathways are all examples of catabolic processes suggests that cells are breaking down complex molecules. There are two possible reasons for this:

- glucose is limited and the cells are using the end products of these catabolic reactions to generate glucose to sustain cellular respiration via alternative entry points to the TCA cycle.
- the cells are using these end products to synthesise other complex molecules e.g., acetyl-CoA is the building block of cholesterol.

Malm et al. (2020) studied the adaptation of HEK293 cell lines from attached to suspension and showed that cholesterol metabolism was significantly upregulated in suspension cell lines compared to adherent. The study presented here didn't assess the impact of adaptation on metabolism in the way Malm et al. (2020) did. However, analysis of GO terms identified in this study demonstrated that cholesterol metabolism as well as membrane rafts, cholesterol rich structures located in the plasma membrane, were significantly enriched when comparing K1 and DXB11 cells adapted from adherent to suspension growth in the presence of 5% FBS.

Given that the overarching theme of this thesis is cellular adhesion and how that might be manipulated for use as a primary recovery tool and the fact that this chapter specifically focuses on the adaptation of K1 and DXB11 cells from attached to suspension growth adhesion pathways and processes were investigated.

Of the three classes of cell junction, it's only anchoring junctions which are involved in cellular adhesion to a surface. For this reason, it's not a surprise that *focal adhesions* were an enriched pathway and GO term while *cadherin binding* was also an enriched GO term. Cadherins tend to be more associated with cell-cell contacts that cell surface contacts. The fact that both these terms are represented in the data generated by this study is good evidence that both cell-surface adhesions as well as intercellular adhesions are disrupted during the adaptation from attached to suspension growth. Occluding

junctions such as tight junctions are more associated with cellular communication and enabling signalling pathways than adhesion. Despite this they were identified as a significantly enriched KEGG pathway involved in adhesion. PSEA analysis was performed on the one hundred differentially expressed proteins which were common to both K1 and DXB11 cell lines. The output of this analysis was filtered for enriched pathways and differentially expressed proteins involved in adhesion (Table 6.8) and the enriched GO terms associated with those proteins (Tables 6.10). This resulted in *heat shock protein 90B1* (HSP90B1), and *cathepsin-D* (both upregulated) while *heat shock protein 90A1*, *poly(ADP-Ribose) Polymerase 1*, *tight junction protein 1*, *the E3 ubiquitin-protein ligase NEDD4*, *laminin subunit alpha 5*, *calpain-2*, *filamin-A* and *cytosolic phospholipase A2* (all downregulated) as being identified as being associated with cellular adhesion.

The data suggests that EMT rather than wound healing is a better model for describing events associated with the adaption of cells from attached to suspension growth as over a third of the one hundred DE proteins identified as being common to both K1 and DXB11 cells following adaption from attached to suspension were present in a published database of genes associated with EMT (35/100). Conversely less than a tenth of the genes present in a published database of genes associated with wound healing were present in the list of one hundred DE proteins identified as being common to both K1 and DXB11 cells following adaption from attached to suspension (6/100).

The process of adaptation from attached to suspension growth could therefore be described as consisting of three steps:

1. Disruption of the interactions between cells and the ECM, i.e., disruption of anchoring junctions e.g., focal adhesions
2. Disruptions of the interactions between neighbouring cells, i.e., disruption of adherens and tight junctions.
3. Reorganisation of the cellular cytoskeleton and plasma membrane.

6.4.3. DE Proteins implicated in the disruption of the Cell/ECM interaction:

HSP90 is a molecular chaperone which has been shown to play a vital role in signalling pathways., It does this by acting as a scaffold thus ensuring that signalling proteins, e.g., kinases, and their target protein are held in close proximity to each other. Multiple authors have reported a link between expression of members the HSP90 family of proteins and focal adhesion kinase (FAK) a key enzyme involved in the turnover of focal adhesions. While the link between FAK expression and cellular adhesion is well understood the evidence regarding the HSP90 family and focal adhesions is conflicting.

Ochel et al. (1999) provided evidence for a link between the expression of HSP90 and FAK. Specifically, they linked increased HSP90 expression with decreased FAK expression in the presence of the antibiotic geldanamycin which is a known tyrosine kinase inhibitor. Le Boeuf et al. (2004) used HUVECs to establish that exposure to VEGF controlled the association of HSP90 to the VEGF receptor VEGFR2 and that this binding event was a prerequisite to the phosphorylation of FAK which is necessary for its catalytic activity. The association of HSP90 and VEGFR2 was inhibited by geldanamycin as was phosphorylation of FAK. Furthermore Rousseau et al. (2000) established by *in vitro* experimentation that the inhibition of FAK phosphorylation by geldanamycin correlated with reduced focal adhesion assembly. What neither of the Ochel or Le Boeuf papers established was which member or members of the HSP90 family was involved in this process. Xiong et al. (2015) established a role for *HSP90B1* in protecting FAK from proteasome-mediated degradation. This ties in with the fact that *HSP90B1* encodes for the chaperone gp96/GRP94. This protein is the ER paralogue of cytosolic HSP90 where it plays a role in the proper folding of Toll like receptors and integrins thus protecting them from degradation by the unfolded protein response. In this study *HSP90B1* expression was found to be upregulated even though focal adhesions as a pathway was downregulated. GO analysis performed as part of this study identified focal adhesion, as a GO term, as being enriched and *HSP90B1* was associated with this term along with five other genes. Filtering of enriched GO terms, for GO terms associated with adhesion, identified focal adhesion, as a GO term, as being enriched but the number of genes

associated with the term was reduced to three including *HSP90B1*. Therefore, the answer to this discrepancy came from analysis of the KEGG pathway map for focal adhesion. This revealed that HSP90 proteins are not a component of this pathway despite their well-documented interactions with FAK. Therefore, it appears if the effect of *HSP90B1* expression on integrin expression were not sufficient to prevent focal adhesion as a pathway being downregulated.

Poly (ADP-ribose) polymerase (PARP1) a nuclear enzyme which plays a role in DNA repair and transcription is known to be controlled by the transcription factor Sp1. Using a rabbit corneal epithelial cell line Zaniolo et al. (2006) showed that binding of the α5β1 integrin to fibronectin started a signalling cascade via the MAPK and PI3K pathways which resulted the hyperphosphorylation of Sp1 and subsequent increase in PARP1 promoter activity. A similar cascade has been observed in a scratch wound model using human endothelial cells (Urbich et al., 2002). Both fibronectin and integrins are key components of focal adhesions. PARP1 is also known to play a role in tissue damage and specifically the induction of cell death in damaged cells by depleting intracellular NAD^+ levels. It's not unreasonable therefore to hypothesise that disruptions to focal adhesions which occur during the adaption from attached to suspension growth inhibit PARP1 activity thus maintaining NAD^+ levels and supporting cell survival. Mathews and Berk (2008) have observed a similar effect in HUVECs. Treating cells with the PARP1 inhibitor PJ34 resulted in the attenuation of the effects of ROS induced cell death and the activation of survival pathways involving VEGFR2, Akt, and BAD phosphorylation.

The proteins *Laminin 5, calpain 2 and filamin A* were all identified as downregulated components of the focal adhesion pathway. Laminins, including laminin 5, are components of the epithelial basement membrane. They function as cell adhesion molecules by binding to specific integrins. Laminin-5 has been shown to bind to integrin α6β4. This binding step is critical for the formation of hemidesmosomes in the skin. Calpain 2 is a calcium dependent thiol protease which cleaves molecules involved in cytoskeletal remodelling and signal transduction. Filamin A is a scaffolding protein which functions by crosslinking actin and the extracellular matrix. It also has a role in integrin receptor signalling.

6.4.4. DE Proteins implicated in the disruption of interactions between cells

Tight junction protein 1, aka Zonula-occludens-1 (ZO-1), and the ubiquitin E3 ligase NEDD4 are key proteins in the formation of tight junctions. Tight junctions are cell-cell contacts which regulate the transport of ions and small molecules between cells. They also play a role in establishing cell polarity, the establishment of which is a prerequisite for the formation of cell-cell contacts. Van Campenhout et al. (2011) described a role for NEDD4 in the ubiquitination of the scaffolding protein Dlg3 and its subsequent recruitment to the apical membrane and tight junction formation. In recent years it has become clear that tight junctions also play a role in the establishment and maintenance of adherens junctions via the activity of ZO-1. It's now known that tight junctions are comprised of several transmembrane proteins, occludin, claudins and junctional adhesion molecules (JAMs), as well as ZO proteins. ZO-1 binds occludin but can also bind to either actin or α-catenin present in adherens junctions. In this way tight junctions play a role in the stabilisation of adherens junctions. Both tight junction protein1 and the ubiquitin E3 ligase were shown to be downregulated upon the adaption of K1 and DXB11 cells from attached to suspension growth.

In addition to the association of focal adhesion formation with increased PARP1 promoter activity as described above there is evidence that inhibiting the expression of PARP1 has the effect of reducing the expression of ICAM1, P-selectin and E selectin (Oshima et al., 2001). These are all molecules involves in cell-cell adhesion as opposed to cell-cell extracellular matrix adhesion. PARP1 was shown to be downregulated upon the adaption of K1 and DXB11 cells from attached to suspension growth.

6.4.5. DE Changes in membrane composition

Membrane (lipid) rafts, which as a GO term were found to be enriched in this study, are cholesterol rich structures located in the plasma membrane which have been shown to be involved in events associated with cellular adhesion. Stahley et al. (2014) demonstrated that desmosomes assembly and disassembly were membrane raft dependent.

A role has been described for HSP90A in the localisation of the heterotrimeric G protein $G\alpha_{12}$ to lipid rafts (Waheed and Jones, 2002). Among the functions of $G\alpha_{12}$ is the activation of serum response factor. Serum response factor has been shown to be essential for actin cytoskeletal organisation and focal adhesion assembly in stem cells (Schratt et al., 2002). The data presented in this chapter demonstrated that HSP90A was down regulated during the adaption of K1 and DXB11 cells from attached to suspension growth.

6.4.6. Analysis of enriched GO terms associated with adhesion

While the role the biological processes and cellular compartments listed in Table 6.10 play in cellular adhesion is clear, the role the GO terms related to molecular function may not be so clear. However, evidence exists that they do in fact play a role in cellular adhesion.

- Aubin, Davy and Soriano (2004) provided evidence for a link between the Smad family of proteins and the ability of transforming growth factor beta (TGFβ) to mediate cellular adhesion in prostate cells.
- Kim et al. (2005) demonstrated a link between histone acetylation and cellular adhesion status. Feichtinger et al. (2016) demonstrated that adapting CHO K1 cells from attached to suspension growth led to an increased number of differentially methylated regions compared to suspension cells which were maintained in culture continuously for six months.
- Haworth and Brackenbury (2019) discussed the role ion channels play in cellular adhesion in their review of ion channels and cancer.
- Mangeat, Roy and Martin (1999) provided evidence that the binding of unfolded proteins plays a role cellular adhesion.
- Ganendren et al. (2006) demonstrated that phosphatidyl phospholipase B activity (Plb1 gene) was necessary for the adhesion of *Cryptococcus neoformans* to human lung epithelial cells during infection.

6.4.1. Impact of correcting for false discovery rates (FDR) on the identification of differentially expressed proteins common to CHO K1 and CHO DXB11 cells following adaptation to suspension growth in media containing serum.

Interrogation of the DE protein list generated by Progenesis QI for Proteomics following application of a retrospective FDR criteria at the DE peptide selection step in Progenesis QI for Proteomics and an assessment of the impact omitting such a criteria had in the identification of DE proteins resulted in a reduction in the overall number of DE proteins in each cell line and in the number of DE proteins common to both CHO K1 and CHO DXB11 cell lines. While the overall picture presented by analysis of enriched KEGG pathways showed that metabolic pathways dominated in both scenarios there was a difference in the type of metabolic pathways which were found to be enriched. In the absence of an FDR filter during peptide selection the metabolism of carbohydrates, lipids and amino acids dominated. The application of an FDR filter during peptide selection resulted in KEGG pathways which were involved in organismal systems rather than metabolism dominating. Additionally, the DE proteins which were found to be part of these pathways support the idea that combating cell stress (*nicotinamide nucleotide transhydrogenase, NMT*) and inhibition of focal adhesion kinase activity and thus cellular adhesion (*phospholipase A2 group IVA, PLA2G4A*) underpin the adaptation of cells from attached to suspension growth.

6.4.1.1. Cell clumping as a survival mechanism during adaptation to suspension growth.

It has been observed that metastasising cancer cells will form clumps as they detach from the ECM (Aceto et al., 2014) and similarly cell clumping was observed during the adaptation of CHO K1 and CHO DXB11 cells from attached to suspension growth in this study. It's also previously been shown that cells which detach from the ECM die (Simpson et al., 2008) via a process which involves both caspase dependent and caspase independent pathways. Hawk and Schafer (2018) and Hawk et al. (2018) demonstrated that both caspase dependent and caspase independent pathways involved increased production of reactive oxygen species (ROS). The function of NNT is to regulate

mitochondrial NADPH levels and oxidative stress by controlling the generation of mitochondrial ROS. Li et al. (2018) and Rao et al. (2020) both demonstrated a role for NNT in reducing mitochondrial ROS levels in gastric cancer cell lines and human aortic endothelial cells respectively. NNT was shown to be upregulated in this study perhaps suggesting a mechanism by which the clumping of CHO K1 and CHO DXB11 cells, during the adaptation from attached to suspension growth, is a protective mechanism which prevents cell death and buys the cell time to adapt to growth away from the ECM support system.

PLA2G4A plays a role in the cleavage of arachidonic acid from membrane phospholipids and was shown to be downregulated following the adaptation of CHO K1 and DXB11 cells from attached to suspension growth. Arachidonic acid is itself a key intermediate in the biosynthesis of eicosanoids. Examples of eicosanoids include eoxins, leukotrienes, lipoxins, prostacyclin, prostaglandins, resolvins, and thromboxanes. In addition to its role in the biosynthesis of eicosanoids, Navarro-Tito et al. (2008) demonstrated a role of arachidonic acid in the autophosphorylation of focal adhesion kinase (FAK) via a pathway which involved Gi/Go proteins and SRC kinases. The role FAK and specifically its phosphorylation status plays in cellular adhesion has been previously mentioned in this thesis. Therefore, it's plausible to suggest that the adaptation of CHO K1 and CHO DXB11 cells from attached to suspension growth was facilitated by the downregulation of PLA2G4A via reduced arachidonic acid levels and the consequent inhibition of FAK autophosphorylation.

6.5. Conclusion

The process of adaptation from attached to suspension growth is known to be a complex one involving multiple proteins and signalling events. Four main conclusions can be drawn from this study presented in this chapter.

1. The evidence presented in this chapter supports the idea that the EMT model is a suitable model for describing the transition from growth as attached cells to growth in suspension.

2. KEGG pathways and proteins known to be adhesion related were identified as being statistically enriched following adaptation of CHO K1 and DXB11 cells from attached to suspension growth in the presence of 5% FBS.
3. The majority of these proteins (8/10) are downregulated during adaption of CHO K1 and DXB11 cells from attached to suspension growth in the presence of 5% FBS.
4. The data suggests that alongside the changes in cellular adhesion, cellular processes involved in metabolism were also enriched. These included catabolic processes such as the breakdown of amino acids and fatty acids as well as processes supporting the cellular response to oxidative stress caused by ROS production.

6.6. DE proteins implicated in enriched KEGG pathways discussed in this chapter

HGNC Symbol	Protein Name	Uniprot ID	Listed In
ACAA2	3-ketoacyl-CoA thiolase, mitochondrial	P42765	Table 6.4
ACADVL	acyl-CoA dehydrogenase very long chain	P49748	Table 6.4
ACAT1	acetyl-CoA acetyltransferase 1	P24752	Table 6.4
ACOX1	acyl-CoA oxidase 1	Q15067	Table 6.4
ADH5	alcohol dehydrogenase 5 (class III), chi polypeptide	P11766	Table 6.4
AKR1A1	aldo-keto reductase family 1 member A1	P14550	Table 6.4
ALDH2	aldehyde dehydrogenase 2 family member	P05091	Table 6.4
ALDH5A1	aldehyde dehydrogenase 5 family member A1	P51649	Table 6.2, Table 6.4
CAD	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase	P27708	Table 6.4
CAPN2	calpain 2	P17655	Table 6.4, Table 6.8
CAPZB	capping actin protein of muscle Z-line subunit beta	P47756	Table 6.4
CTSD	cathepsin D	P07339	Table 6.4, Table 6.8
DNM2	dynamin 2	P50570	Table 6.4
FASN	fatty acid synthase	P49327	Table 6.4
FLOT1	flotillin 1	O75955	Table 6.4
FLNA	filamin A	P21333	Table 6.8

HGNC Symbol	Protein Name	Uniprot ID	Listed In
GANAB	glucosidase II alpha subunit	Q14697	Table 6.4
GLUD1	glutamate dehydrogenase 1	P00367	Table 6.4
GOT2	glutamic-oxaloacetic transaminase 2	P00505	Table 6.4
GSTP1	G1 to S phase transition 1	P15170	Table 6.4
GUSB	glucuronidase beta	P08236	Table 6.4
GYS1	glycogen synthase 1	P13807	Table 6.4
HADHB	hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex	P55084	Table 6.4
HSD17B4	hydroxysteroid 17-beta dehydrogenase 4	P51659	Table 6.4
HSP90AA1	heat shock protein 90 alpha family class A member 1	P07900	Table 6.4, Table 6.8
HSP90B1	heat shock protein 90 beta family member 1	P14625	Table 6.4, Table 6.8
IDH2	isocitrate dehydrogenase (NADP(+)) 2	P48735	Table 6.4
KIF5B	kinesin family member 5B	P33176	Table 6.4
LAMA5	laminin subunit alpha 5	O15230	Table 6.4, Table 6.8
LAP3	leucine aminopeptidase 3	P28838	Table 6.4
LMAN2	lectin, mannose binding 2	Q12907	Table 6.4
NEDD4	E3 ubiquitin protein ligase NEDD4	P46934	Table 6.2, Table 6.4 and Table 6.8
NNT	nicotinamide nucleotide transhydrogenase	Q13423	Table 6.8
NPLOC4	NPL4 homolog, ubiquitin recognition factor	Q8TAT6	Table 6.4

HGNC Symbol	Protein Name	Uniprot ID	Listed In
PARP1	poly(ADP-ribose) polymerase 1	P09874	Table 6.4, Table 6.8
PCCB	propionyl-CoA carboxylase subunit beta	P05166	Table 6.4
PDIA6	protein disulfide isomerase family A member 6	Q15084	Table 6.4
PLA2G4A	phospholipase A2 group IVA	P47712	Table 6.8
PRDX1	peroxiredoxin 1	Q06830	Table 6.4
PYGL	glycogen phosphorylase L	P06737	Table 6.4
RRBP1	ribosome binding protein 1	Q9P2E9	Table 6.4
SCP2	sterol carrier protein 2	P22307	Table 6.4
SEC24C	SEC24 homolog C, COPII coat complex component	P53992	Table 6.4
SNX1	sorting nexin 1	Q13596	Table 6.4
TJP1	tight junction protein 1	Q07157	Table 6.2, Table 6.4 and Table 6.8
TXNDC5	thioredoxin domain containing 5	Q8NBS9	Table 6.4
UGP2	UDP-glucose pyrophosphorylase 2	Q16851	Table 6.4

Table 6.12: **Protein names associated with the HGNC symbols listed in Table 6.4, and Table 6.8.**

CHAPTER 7

A proteomic comparison of CHO K1 and DXB11 cells following adaption from attached to suspension growth in serum free media

7.1. Introduction

This thesis set out to develop a flocculating CHO cell line as an alternative approach to the use of commercial flocculants during primary recovery operations at a commercial scale. Initially, the cloning and expression of exogenous proteins, which were considered to have the potential to drive cellular flocculation of CHO cells, was investigated, but this approach was unsuccessful. Therefore, alternative approaches were investigated. Anecdotally, the adaptation of CHO cells to growth in serum free media has been correlated with a clumping phenotype. This raised the question of whether an investigation into serum removal and cellular clumping could identify endogenous CHO proteins that could be engineered to drive cellular flocculation? Consequently, a proteomic study investigating changes in the proteome of CHO K1 cells, which had been pre-adapted to suspension, under different conditions was performed and discussed in chapter five. This proteomic study compared growth in attached culture, in the presence and absence of serum, with growth in suspension culture, also in the presence and absence of serum. However, this study relied upon CHO K1 cells pre-adapted to suspension culture. This raised the question, was the comparison an appropriate one? After all, the molecular and proteomic changes which should result from adaptation to serum free suspension growth had already occurred in these cells, so why would simply adding serum and culturing them as attached cells reverse these molecular and proteomic changes?

The study described in this chapter was designed to identify endogenous CHO proteins, whose expression changed as the serum concentration in the media was reduced, using a differential expression proteomic approach. To answer this question, a vial of attached CHO K1 cells was purchased from the European Collection of Authenticated Cell Culture (ECACC), initially cultured in 5% serum and then, over several months, adapted to serum free suspension growth. A vial of CHO DXB11 was obtained from Prof. Larry Chasin of Columbia University in New York as a second comparison. These were similarly cultured in 5% serum and then adapted to serum free suspension growth over several months. Samples were taken at different serum concentrations during the adaptation process,

thus allowing for a temporal profile of changes in the proteome of both cell lines to be constructed. This chapter discusses the adaption of these cells from growing as a suspension culture in media supplemented with 5% serum to serum free media. The adaptation of these cells, from attached to suspension growth in media supplemented with serum, was discussed in chapter six.

While various groups have studied the proteome of suspension adapted CHO cells a comparison of the changes which occurred in the proteome of different CHO cell lines which were sampled while undergoing adaptation from attached to suspension culture and from serum-dependency to serum-free has not been described. Examples of these studies include Walther et al. (2016) who investigated differences in the expression of specific membrane proteins, the Borth group at BOKU in Vienna performed mRNA and miRNA profiling of CHO K1 suspension cultures (Bort et al., 2012) and studied changes in the transcriptome of CHO cells post adaption to suspension growth (Shridhar et al., 2017). Lee et al. (2016) used an RNA-seq approach followed by CRISPR engineering to accelerate the adaptation of CHO cells from attached to suspension growth. Xu et al. (2017) and Malm et al. (2020) performed a comparative analysis of 3 different lineages of suspension adapted CHO cells and six different HEK293 cell lines respectively.

7.2. Experimental Design and Sample Preparation

The process of adapting these cell lines from attached to suspension growth in media supplemented with serum was described in chapter 6. This chapter describes the adaption process which followed to adapt these suspension cells from growing in media supplemented with serum to growth in serum free media.

Suspension adapted cultures of CHO K1 and CHO DXB11 cells growing in media containing 5% FBS (Section 6.2) were counted daily and passaged every three days until culture viability was above 90%. Specific growth rate (Table 7.1) was used to determine when it was appropriate to reduce the serum concentration. A specific growth rate (μ) of 0.7/day corresponded to a doubling time of 1 day. At this point cells were considered adapted to that serum concentration and the amount of serum in the media was

reduced. The serum concentration was then reduced in a stepwise fashion from 5% to 1% to 0.5% and finally to 0% serum. Cells were taken at each serum concentration, pelleted by centrifugation, washed with ice cold PBS, and stored at -80^{o}C until required.

To identify proteins, which differed in their expression levels between the experimental conditions under investigation, a comparative analysis was performed using a label-free quantitative differential expression LC-MS approach. The approach taken utilised four biological replicates of each condition as outlined in Figure 7.1.

An identical approach to chapter six was taken as regards sample preparation and data analysis using the FASP protocol developed by Coleman et al. (2017) (Section 2.6.3.7). To reduce variability and retention time drift, all timepoints corresponding to each cell line were prepared and analysed on the mass spectrophotometer in a single analysis and the "Two-Sample Analysis Approach'' described by Kelly (2016) was adopted.

Following alignment, samples were split into experimental groups and identified peptides were filtered based on an ANOVA p value of ≤ 0.05 between experimental groups. A mascot generic file (.mgf) was generated for each replicate and exported to Proteome Discoverer. Following peptide and protein identification by Proteome Discoverer, using a proteogenomic draft annotation of the CHO genome Li et al. (2018) (Section 2.6.5), the data was saved as a (.pepxml) file and reimported into Progenesis QI for Proteomics. Criteria of an ANOVA p value of ≤ 0.05, a relative minimum fold change of +/-1.5 between experimental groups and ≥2 peptides contributing to protein identifications was used to identify proteins whose expression profile differed between experimental groups.

All subsequent data manipulation, protein naming, and statistical analysis was performed in the R programming environment (Section 2.6.6).

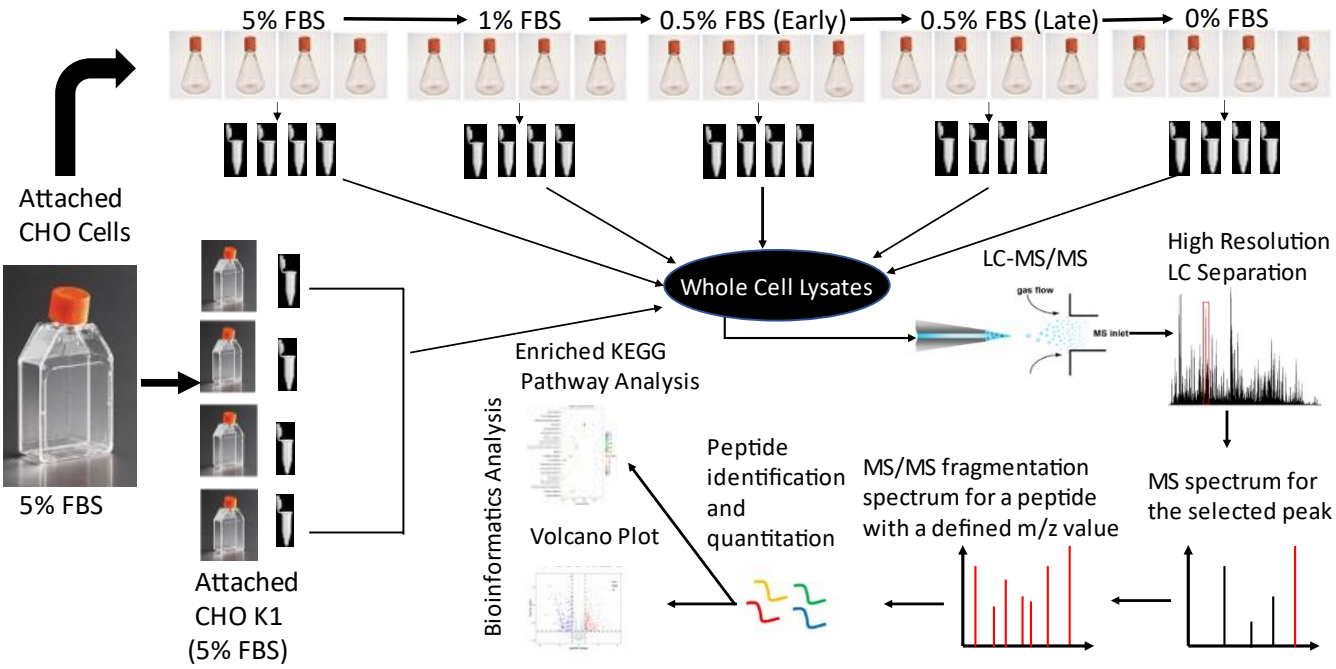


Figure 7.1: **Workflow followed when preparing CHO K1 samples adapted to serum free suspension growth for LC-MS analysis. An identical approach was followed for DXB11 cells adapted to serum free suspension growth.**

7.3. Results

Cells were adapted from 5% FBS to 0% FBS in a stepwise fashion. Table 7.1 summarises the passage number at each step and whether clumping was observed in the culture or not.

K1 cells took longer to adapt to media supplemented with only 0.5% serum than DXB11 cells (seventeen passages for K1 cells against ten passages for DXB11 cells). However once adapted to growth in media containing 0.5% FBS, K1 cells adapted to serum-free conditions much quicker than DXB11 cells (137 days in total for K1 and 261 days in total for DXB11).

Cell pellets corresponding to each of the serum concentrations described in Table 7.1 were processed for MS analysis as per Sections 2.6.4 and 2.6.5.The resulting DE protein lists were processed as described in chapter 6:

Serum Conc	CHO K1				CHO DXB11			
	Passage Number	**Clumping Observed**	**Specific Growth Rate[c]**	**No of Days (Cumulative)**	**Passage Number**	**Clumping Observed**	**Specific Growth Rate[c]**	**No of Days (Cumulative)**
5%	P1	Yes	0.3 (Day 2)	0	P1	Yes	0.4 (Day 2)	0
1%	P8	Yes	0.5 (Day 2)	23	P6	Yes	0.4 (Day 2)	17
0.5%	P17	Yes	0.4 (Day 3)	74	P10	Yes	0.7 (Day 2)	33
0% (Start)	P19	Yes	0.5 (Day 3)	81	P35	Yes	0.6 (Day 2)	181
0% a (Finish)	P34	No	0.7 (Day 3)	137	P40	Yes	0.7 (Day 2)	207
0% b (Start)	NA				P40	Yes	0.7 (Day 2)	207
0% (Finish)					P59	No	0.8 (Day2)	261

a) Corresponds to the point when Covid 19 forced a cessation of all lab activities and cells were frozen and stored @ -70^{o}C.

b) Corresponds to the point when cells were revived once we restarted lab activities.

c) Units for specific growth rate are day^{-1}.

Table 7.1: Summary of the adaptation process. CHO DXB11 and K1 cells, which had been grown in T175 flasks, were trypsinised and used to seed a separate 250mL shake flask for each cell line. These flasks were passed multiple times before being used to seed four replicate 250mL shake flasks for each cell line. Media for CHO K1 was Balan CD growth formula A supplemented with serum and L-Glutamine. Media for CHO DXB11 was Balan CD growth formula A supplemented with serum, Hypoxanthine, Thymidine and L-Glutamine. The specific growth rate was reported for the day on which the specific growth rate was highest.

7.3.1. Assessment of replicate clusters

Two approaches were taken to the clustering analysis. These were principal component analysis (Section 2.6.8) (PCA) which is part of the Progenesis QI for Proteomics workflow, and the use of heat maps (Section 2.6.7) based on the Euclidean distance.

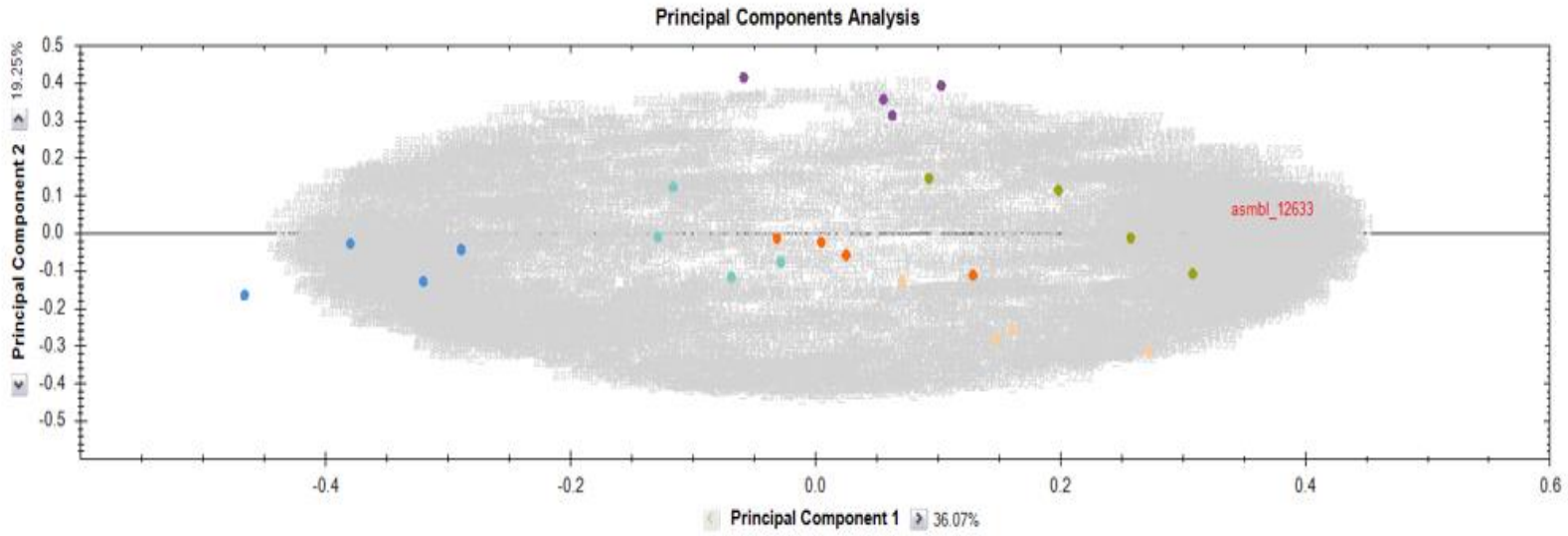


Figure 7.2: Principal Component Analysis clustering of the various serum concentration analysed during the adaptation of K1 cells from attached growth in media supplemented with serum to suspension growth in serum free media (N=4). PCA analysis was performed by the Progenesis QI for Proteomics software after identification of DE proteins as a quality control step to assess the clustering of replicate samples. The different coloured dots represent different serum concentrations. *Blue dots (5% attached), purple dots (5% suspension), orange dots (1% suspension), green dots (0.5% early suspension), beige dots (0.5% late suspension) and turquoise dots (0% suspension). The grey cloud represents individual differentially expressed proteins.*

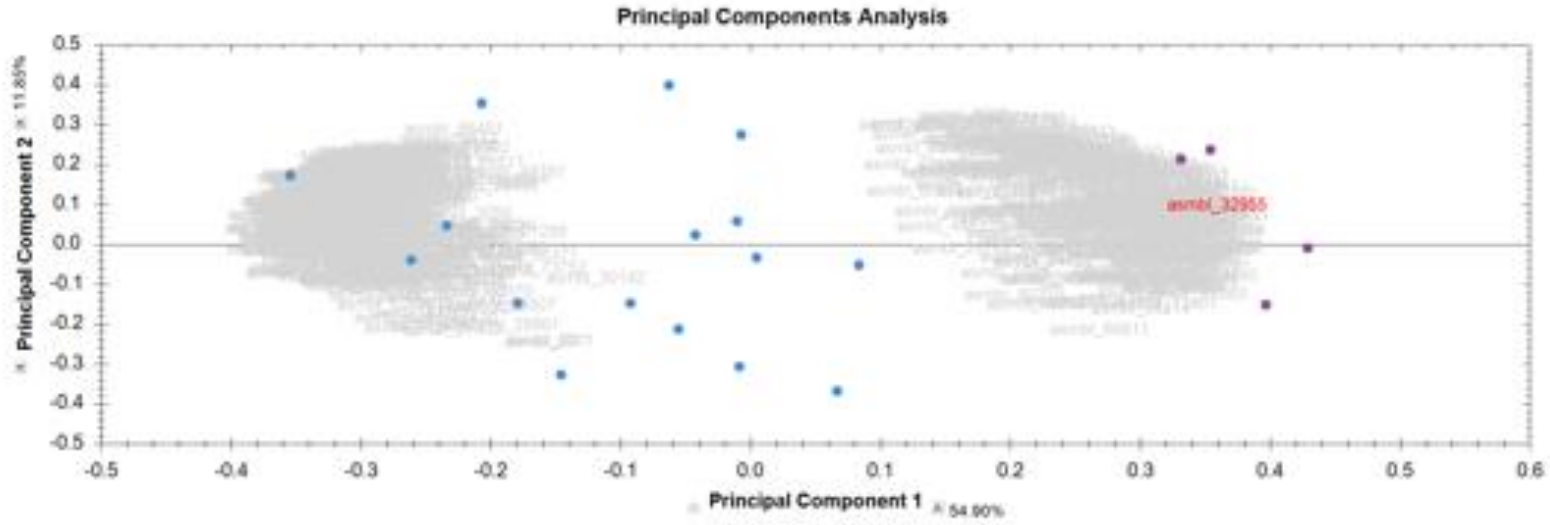


Figure 7.3: **Principal Component Analysis clustering of the suspension (serum) v suspension (no serum) comparison of K1 cells (N=4). PCA analysis was performed by the Progenesis QI for Proteomics software after identification of DE proteins as a quality control step to assess the clustering of replicate samples.** *Blue dots represent cells grown in media containing serum (5%, 1%, 0.5% early and 0.5% late). e. Purple dots represent cells grown in serum free media. The grey clouds represent individual proteins found to be differentially expressed in each condition.*

Analysis of Figure 7.2 suggests that, except for the CHO K1 cells grown in suspension with 5% FBS (purple dots) or 0.5% late suspension (beige dots), the variation in the data can be described by principal component 1. In comparison the variation in the data for CHO K1 cells grown in suspension with 5% FBS or 0.5% FBS late can be described by both principal component 1 and 2. The fact that the individual replicates for a particular serum concentration did not always cluster together (e.g., green and beige dots) was attributed to the fact that these were biological replicates (individual flasks) rather than technical replicates (multiple samples taken from a single flask) undergoing an adaptation process and a degree of variability in the response of each flask is to be expected. Examination of the clustering pattern in Figure 7.2 it could be suggested that attached CHO K1 cells grown in media supplemented with 5% serum (blue dots) have differences in their proteome compared to suspension adapted CHO K1 cells grown in media with 5% serum (purple dots) which are also different, on a proteomic level, to suspension adapted CHO K1 cells grown in serum free media (purple dots). However, based on Figure 7.2 it would be difficult to attribute differences in an individual DE protein's expression profile to the different growth conditions based on principal components 1 and 2. Therefore, the PCA analysis was repeated by excluding the 5% attached CHO K1 cells and grouping the suspension adapted CHO K1 cells into two experimental groups based on whether serum was present in the media or not (Figure 7.3).

Analysis of Figure 7.3 supports the theory that adaption of CHO K1 cells to growth in serum free media results in distinct proteomic changes as CHO K1 cells grown in the absence of serum cluster away from CHO K1 cells grown in the presence of serum and two distinct groups of DE proteins can be observed. While the variation in the serum free CHO K1 cells (turquoise dots) is primarily described by principal component 2 the variation in the CHO K1 cells grown in the presence of serum is described by both principal component 1 and 2. The lack of clustering among the CHO K1 cells grown in the presence of serum can be explained by the fact that this experimental condition encompasses four different sampling points and due to the experimental setup chosen for the analysis, identification of the individual sampling points was not possible.

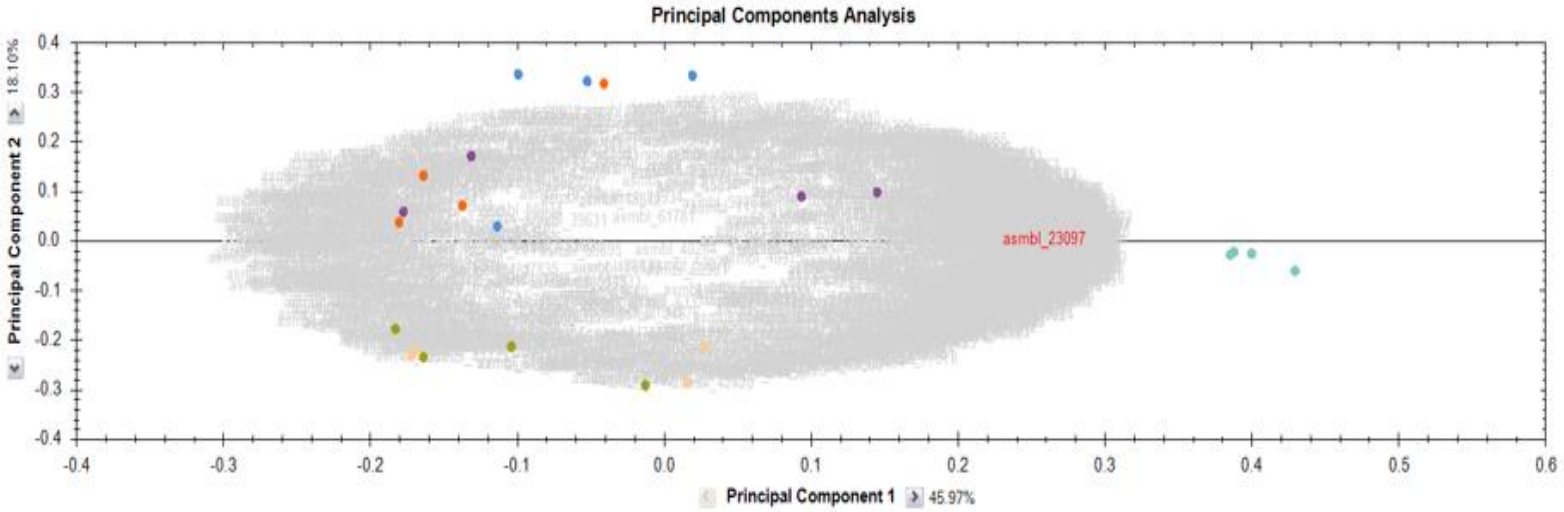


Figure 7.4: **Principal Component Analysis clustering of the various serum concentration analysed during the adaptation of DXB11 cells from attached growth in media supplemented with serum to suspension growth in serum free media (N=4). PCA analysis was performed by the Progenesis QI for Proteomics software after identification of DE proteins as a quality control step to assess the clustering of replicate samples. The different coloured dots represent different serum concentrations.** *Blue dots (5% attached), purple dots (5% suspension), orange dots (1% suspension), green dots (0.5% early suspension), beige dots (0.5% late suspension) and turquoise dots (0% suspension). The grey cloud represents individual differentially expressed proteins.*

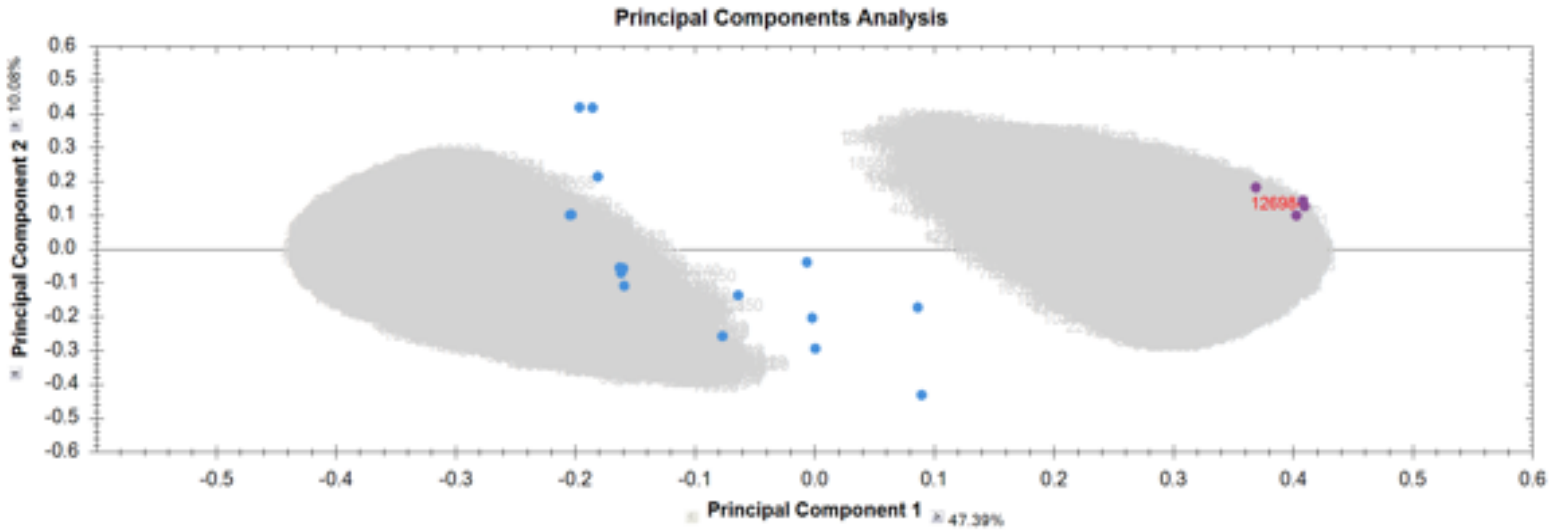


Figure 7.5: **Principal Component Analysis clustering of the suspension (serum) v suspension (no serum) comparison of DXB11 cells (N=4). PCA analysis was performed by the Progenesis QI for Proteomics software after identification of DE proteins as a quality control step to assess the clustering of replicate samples.** *Blue dots represent cells grown in media containing serum (5%, 1%, 0.5% early and 0.5% late). Purple dots represent cells grown in serum free media. The grey clouds represent individual proteins found to be differentially expressed in each condition.*

The PCA analysis of DXB11 cells during the adaptation from attached growth in media supplemented with serum to suspension growth in serum free media (Figures 7.4 and 7.5) showed that except for 2 replicates there was good clustering of replicates within the 5% attached, 1% suspension. 0.5% suspension and 0% suspension adapted DXB11 cells. The exceptions were a one 5% attached and one 1% suspension replicate which swapped groups. It's not clear what is the reason behind this. The replicates of the 5% suspension adapted DXB11 cells didn't group particularly well. Two of the replicates clustered towards attached and suspension adapted cells grown in the presence of serum while the other two clustered towards suspension adapted cells grown in serum free media. However, this may be visual evidence of the adaptation process occurring. The variation in the 5% suspension adapted, both 0.5% suspension adapted time points and the 0% suspension adapted DXB11 cells is described primarily by principal component one while the variation associated with the 5% attached and 1% suspension adapted DXB11 cells is described by both principal components one and two.

Analysis of Figure 7.4 suggested that attached CHO DXB11 cells grown in media supplemented with 5% serum (blue dots) have differences in their proteome compared to suspension adapted CHO DXB11 cells grown in media with 0.5% serum (green and beige dots) which are also different, on a proteomic level, to suspension adapted CHO DXB11 cells grown in serum free media (turquoise dots). Comparison of Figures 7.2 and 7.4 also suggests that there is a greater degree of variability between the proteomes of 5% attached and 5% suspension DXB11 cells compared to the same CHO K1 cells. This observation is based on the greater degree of separation between the respective groups along principal component one. It's difficult to know whether this is a true difference or is due to the fact that the CHO DXB11 cells were allowed to adapt for a longer period of time (Table 7.1). However, similar to Figure 7.2 it would be difficult to attribute differences in an individual DE protein's expression profile to the different growth conditions based on the profile presented in Figure 7.4. Therefore, the PCA analysis was repeated by excluding the 5% attached CHO DXB11 cells and grouping the suspension adapted CHO DXB11 cells into two experimental groups based on whether serum was present in the media or not (Figure 7.5).

Analysis of Figure 7.5 supports the theory that adaption of CHO DXB11 cells to growth in serum free media results in distinct proteomic changes as CHO DXB11 cells grown in

the absence of serum cluster away from CHO DXB11 cells grown in the presence of serum and two distinct groups of DE proteins can be observed. While the variation in the serum free CHO DXB11 cells (purple dots) is primarily described by principal component 1 the variation in the CHO DXB11 cells grown in the presence of serum is described by both principal component one and two. The lack of clustering among the CHO DXB11 cells grown in the absence of serum can be explained by the fact that this experimental condition encompasses four different sampling points and due to the experimental setup chosen for the analysis identification of the individual sampling points was not possible.

To further examine the clustering of the different experimental groups heatmaps based on the Euclidean distance were generated. All four replicates from each experimental group (attached cells 5% FBS, suspension cells 5% FBS, suspension cells 1% FBS, suspension cells 0.5% FBS early, suspension cells, 0.5% FBS late and suspension cells 0% FBS) were analysed (Figure 7.6 and Figure 7.7).

Figures 7.6 and Figure7.7 demonstrated that replicates fall into several clusters. DXB11 cells can be subdivided as follows. Suspension cells grown in media supplemented 5% serum don’t cluster and form two distinct populations. Two of 5% suspension replicates cluster with the 5% attached and 1% suspension cells while the other two cluster with the 0% suspension cells. The 5% attached cells cluster with the 1% suspensions cells but are intermingled as opposed to clustering together, while cells grown in media supplemented with 0.5% serum or serum free media cluster tightly together. CHO-K1 cell samples don't show the same pattern. The individual replicates cluster tightly together.

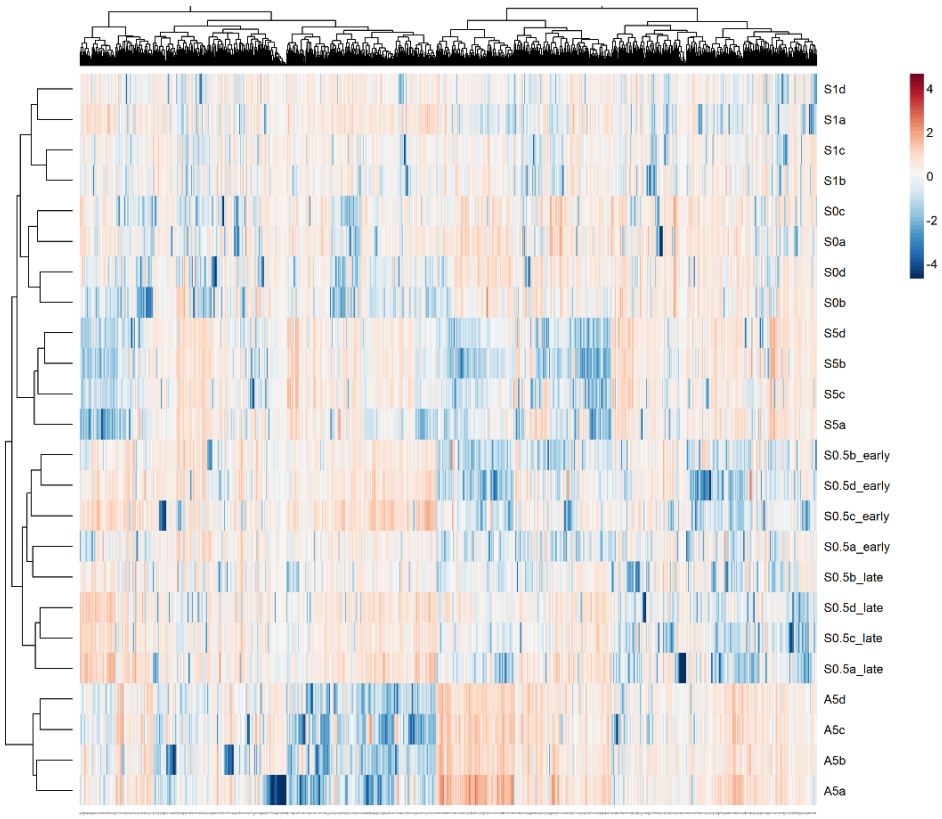


Figure 7.6 Clustering of CHO K1 cells based on the Euclidean distance between samples. All DE proteins with a p ≤ 0.05 and fold change ≥ 1.5 (identified by Progenesis QI for Proteomics) are displayed. *A5 = attached cells grown in the presence of 5% serum, S5 = cells grown in 5% serum as suspension cells, S1 = cells grown in 1% serum as suspension cells, S0.5 = cells grown in 0.5% serum as suspension cells, S0 = cells gown in the absence of serum.*

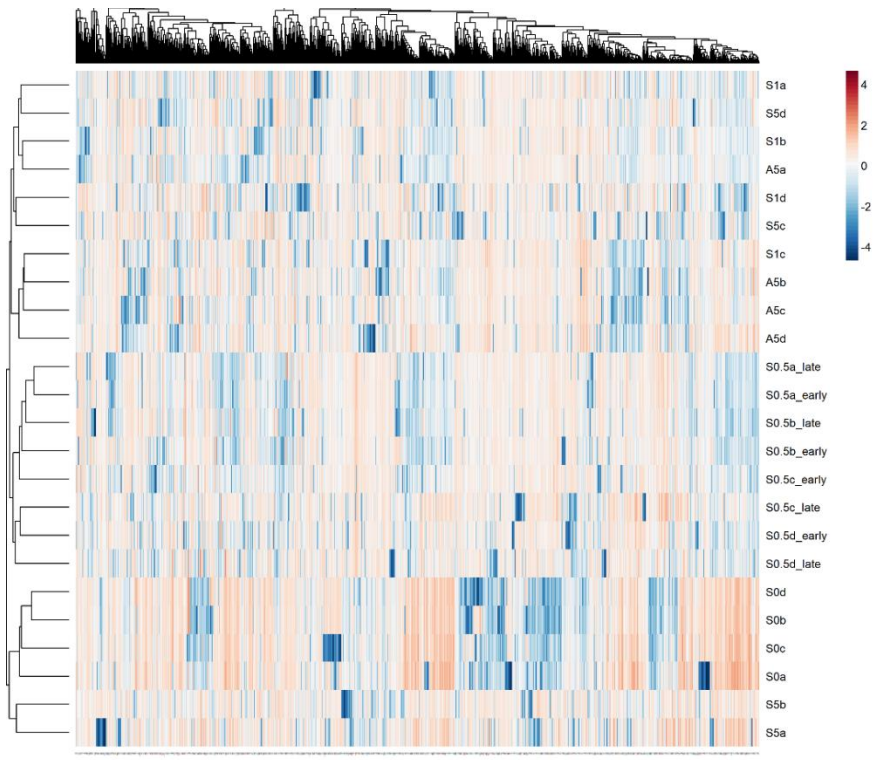


Figure 7.7 Clustering of DXB11 cells based on the Euclidean distance between samples. All DE proteins with a p ≤ 0.05 and fold change ≥ 1.5 (identified by Progenesis QI for Proteomics) are displayed.

A5 = attached cells grown in the presence of 5% serum, S5 = cells grown in 5% serum as suspension cells, S1 = cells grown in 1% serum as suspension cells, S0.5 = cells grown in 0.5% serum as suspension cells, S0 = cells gown in the absence of serum.

Having determined, based on the PCA and heatmap analysis, that comparing suspension adapted cells grown in media containing serum with suspension adapted cells grown in serum free media gave a clearer picture all subsequent analysis was performed on that basis.

It is also possible that the clustering pattern observed in Figures 7.2, 7.4, 7.6 and 7.7 is due a lack of reproducibility during cell lysis and sample preparation. As discussed in chapter 5 if the total number of proteins detected by the mass spec differs between replicates of the same sample or if a particular protein is not detected in all replicates of this will increase variability and could potentially lead to a protein being identified as differentially expressed in error. To investigate this, the total number of proteins detected in each replicate was quantified using Proteome Discoverer. These numbers

are displayed in Table 7.2. It's also worth commenting that samples were analysed on a single occasion as biological replicates. Reproducibility could have been increased by also performing technical replicates, i.e., performing multiple injections from each MS vial and selecting only proteins which were detected in all technical replicates for each biological replicate.

Experimental Group	Replicate	Total No. of Proteins	Average No. of Proteins	No. of Common Proteins	No. of DE Proteins
K1 Suspension (5% FBS)	S5a	3261	3323	2471	FC = 1.5
	S5b	3480			
	S5c	3043			
	S5d	3506			
K1 Suspension (1% FBS)	S1a	3206	3147	2359	
	S1b	3181			
	S1c	3194			
	S1d	3005			
K1 Suspension (0.5% FBS early)	S0.5 early a	3076	3317	2432	
	S0.5 early b	3476			
	S0.5 early c	3252			
	S0.5 early d	3464			
K1 Suspension (0.5% FBS late)	S0.5 late a	3635	3463	2594	
	S0.5 late b	3318			
	S0.5 late c	3298			
	S0.5 late d	3599			
K1 Suspension (0% FBS)	S0a	2952	3252	2374	
	S0b	3273			
	S0c	3394			
	S0d	3388			
DXB11 Suspension (5% FBS)	S5a	4287	3312	1714	FC = 1.5
	S5b	4201			
	S5c	1982			
	S5d	2776			
DXB11 Suspension (1% FBS)	S1a	2695	2938	1657	
	S1b	3362			
	S1c	3710			
	S1d	1986			
DXB11 Suspension (0.5% FBS early)	S0.5 early a	3607	3566	2471	
	S0.5 early b	2921			
	S0.5 early c	3715			
	S0.5 early d	4022			
DXB11 Suspension (0.5% FBS late)	S0.5 late a	3534	3665	2684	
	S0.5 late b	3984			
	S0.5 late c	3575			
	S0.5 late d	3568			
DXB11 Suspension (0% FBS)	S0a	4124	3989	3172	
	S0b	4082			
	S0c	3665			
	S0d	4084			

Table 7.2: **Total numbers of proteins detected in all replicates. MS raw data files were imported directly into Proteome Discoverer and proteins detected as per Section 2.6.5. The number of common proteins column recorded the number of proteins detected in all four replicates.**

In excess of three thousand proteins were identified in each replicate for all K1 experimental groups. However, the DXB11 experimental groups showed greater variability, in particular the 5% FBS and 1% FBS concentrations. The number of proteins detected in the DXB11 5% FBS and 1% FBS experimental groups ranged from one thousand nine hundred and eighty six to four thousand two hundred and eighty seven. The impact of this variability is that not all proteins were detected in each replicate. In order to assess the impact of this the proteins common to all replicates of each experimental group were identified and compared to the DE proteins identified as playing a role in the KEGG pathways (Section 7.3.3). This enabled the identification of six DE proteins present in these pathways which were not common to all replicates. These proteins are listed in Table 7.3.

KEGG Path	HGNC Symbol	Protein Name	Uniprot ID	Direction
Terpenoid backbone biosynthesis	HMGCS2	3-hydroxy-3-methylglutaryl-CoA synthase 2	P54868	Down regulated
Ribosome	MRPS9	mitochondrial ribosomal protein S9	P82933	Down regulated
Purine metabolism	PFAS	phosphoribosylformylglycinamidine synthase	O15067	Up regulated
Cholesterol metabolism	SCARB1	scavenger receptor class B member 1	Q8WTV0	Down regulated
	VAPA	VAMP associated protein A	Q9P0L0	
Protein Processing in the endoplasmic reticulum	SIL1	SIL1 nucleotide exchange factor	Q9H173	Down regulated

Table 7.3: **DE proteins detected in enriched KEGG paths which were not detected all replicates of each serum concentration in both cell lines.**

7.3.2. Identification of differentially expressed proteins with a common expression profile in CHO K1 and DXB11 cells after adaptation to growth in serum free media

One thousand one hundred and eighty DE proteins were detected in CHO DXB11 cells compared to one thousand six hundred and nine in CHO K1 cells following adaptation to suspension growth in serum free media. Once the appropriate filters had been applied, max fold change as calculated by Progenesis QI for Proteomics ≥1.5, an p value ≤ 0.05, a false discovery rate of 1% and at least two unique peptides contributing to protein identification, five hundred and two of these DE proteins were identified as being unique to CHO K1 cells. (Appendix 20). The same filtering step identified one thousand one hundred and seven DE proteins as unique to DXB11 cells (Appendix 21).

Table 7.4 lists the top 5 proteins in either cell line based on:

a. Absolute fold change (absolute FC) between the two experimental groups
b. p value

Cell Line	Protein Name	HGNC Symbol	p Value	Max FC[1]	Log_2 (Absolute FC)[2]	Present in all serum containing reps	Present in all serum free reps
Ranked by absolute FC							
K1	cytochrome P450 family 4 subfamily B member 1	CYP4B1	3.45E-04	75.06	-6.230	Not detected in S5a, S5b, S5c, S5d, S1a, S1b, S1c, S1d, S0.5a_late, S0.5b_late, S0.5c_late, S0.5d_late	No
	S100 calcium binding protein-A5	S100A5	3.10E-05	16.602	-4.053	Not detected in S1a, S1c, S0.5a_early, S0.5d_early, S0.5a_late, S0.5c_late, S0.5d_late	No

K1	Collagen type III alpha 1 chain	COL3A1	5.97E-05	11.174	-3.482	Not detected in S1d,	No
	C-X-C motif chemokine ligand 12	CXCL12	0.0002	10.425	-3.382	Not detected in S5a, S5b, S5c	No
	Lumican	LUM	7.65E-05	8.239	-3.042	Not detected in S5a, S5d, S1a, S1b, S1d, S0.5_late, S0.5c_late	No
DXB11	SPARC	SPARC	4.35E-09	171.801	-7.158	Not detected in S5b, S1c, S0.5a_early	Not detected in S0b
	Aldo-keto reductase family member B7	AKR1B7	3.76E-05	102.476	-6.059	Not detected in S5b, S1b, S1c, S0.5d_late	Not detected in S0a, S0b, S0c
	Tropomyosin 2	TPM2	7.34E-07	33.3200	4.769	Yes	Yes

	FERM, ARHGEF and pleckstrin domain-containing protein 1[3]	FARP1	5.97E-05	33.283	-4.718	Not detected in S5b, S5c, S5d, S1c, S0.5c_early, S0.5b_late, S0.5c_late, S0.5d_late	Not detected in S0a, S0b, S0c
	Semaphorin 3E	SEMA3E	0.0021	65.187	4.705	No	Yes
Ranked by p value (order of significance)							
K1	Annexin	ANXA1	1.51E-07	2.180	1.125	Yes	Yes
	Endoplasmic reticulum-Golgi intermediate compartment protein 2	ERGIC2	1.65E-07	1.820	-0.864	Yes	Not detected in S0a, S0d,
	Postaglandin reductase 1	PTGR1	1.65E-07	1.635	0.710	Yes	Yes
	Endoplasmic reticulum-Golgi intermediate compartment protein 3	ERGIC3	1.66E-07	1.614	-0.690	Yes	Not detected in S0a, S0b, S0d,
	Collagen type VI alpha 1 chain	COL6A1	2.54E-07	7.565	-2.919	Yes	Yes

DXB11	40S ribosomal protein S18	RPS18	7.11E-11	8.335	2.999	Yes	Yes
	Uridine monophosphate synthetase	UMPS	2.02E-10	5.551	1.911	Yes	Yes
	Amyloid beta precursor protein	APP	1.98E-09	6.078	1.804	Not detected in S5b, S5d, S1a, S1c, S0.5a_early	Yes
	Lipoprotein lipase	LPL	2.46E-09	7.464	1.258	Yes	Yes
	Peptidyl-glycine alpha-amidating monooxygenase	PAM	2.68E-09	11.642	2.150	Not detected in S5b, S5c, S5d, S1a, S1b	Yes

[1] Max FC is the proteins fold change value as calculated by Progenesis QI for Proteomics.

[2] Log2 Absolute FC allows a direction to be placed on the fold change. A value > 0 indicated the protein is upregulated in the comparison while a value < 0 indicates the protein is downregulated. It was calculated by dividing the proteins mean abundance in cells grown in serum free media by its mean abundance in cells grown in media containing serum.

[3] Although annotated in UniprotKB this protein has not been reviewed and was not included in further analysis.

Table 7.4: The five most significant DE proteins detected in CHO K1 and CHO DXB11 cell lines as ranked by max fold change and statistical significance.

While the subsequent analysis could be confined to these proteins a different approach was taken. It was decided to focus on DE proteins which were common to both cell lines and displayed a common expression profile in response to the removal of serum from the media. Bearing in mind what the experiment was designed to identify, proteins implicated in the adaptation of cells from growth in media containing serum to growth in serum free media, it seemed logical to focus on these proteins rather than those with the lowest p value or highest fold change.

Two different techniques, volcano plots and upset plots, were used to examine all DE proteins detected in both K1 and DXB11 cells.

The use of volcano plots (Figures 7.8 and 7.9) aided the interpretation of whether the comparison resulted in DE proteins that were upregulated or downregulated. Each point on a volcano plot represents the fold change between two experimental groups plotted against the level of statistical difference.

The fold change in an individual proteins' expression was defined as

$$\frac{average\ normalised\ protein\ abundance\ in\ cells\ adapted\ to\ serum\ free\ media}{average\ normalised\ protein\ abundance\ in\ cells\ grown\ in\ media\ containing\ serum}$$

Volcano plots of the DE proteins detected in CHO K1 and CHO DXB11 cells are presented in Figures 7.8 and 7.9. Volcano plots of statistical significance ($-\log_{10}$Pvalue) versus fold change ($\log_2$ FC) provide an overall impression of whether and how many proteins are up or down regulated within a dataset. By $\log_2$ transforming the data before plotting the plot can be centred around zero. DE proteins with a Log_2 FC > 0 are considered upregulated (these fall on the right hand side of Figures 7.8 and 7.9. DE proteins on the left-hand side of Figures 7.8 and 7.9 have a Log_2 FC < 0 and are downregulated. Similarly, by log transforming the p values associated with each data point ($-\text{Log}_{10}$) highly significant (v small p values) proteins are positioned towards the top of the plot and while those clustered towards the bottom of the plot are less significant.

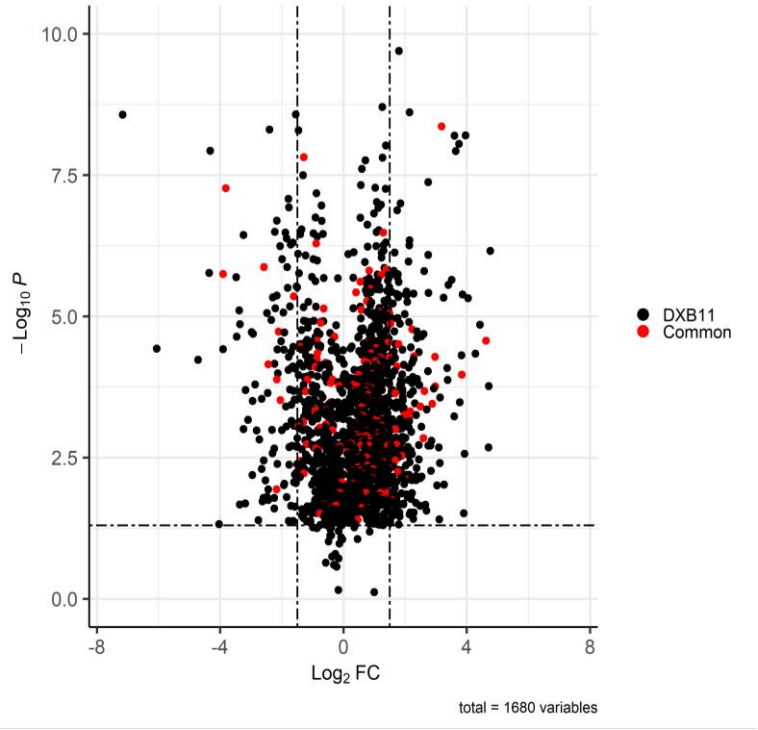


Figure 7.8: All one thousand six hundred and eighty unique DE proteins detected in CHO DXB11 cells following adaptation from suspension growth in media containing (5%) serum to serum free (0%) media. Red dots represent the one hundred and sixty-five DE proteins common to both CHO K1 and DXB11 cells. The vertical dashed lines in Figure 7.7 represent a fold change of ± 1.5 and the horizontal dashed line represents $log_{10}(0.05)$.

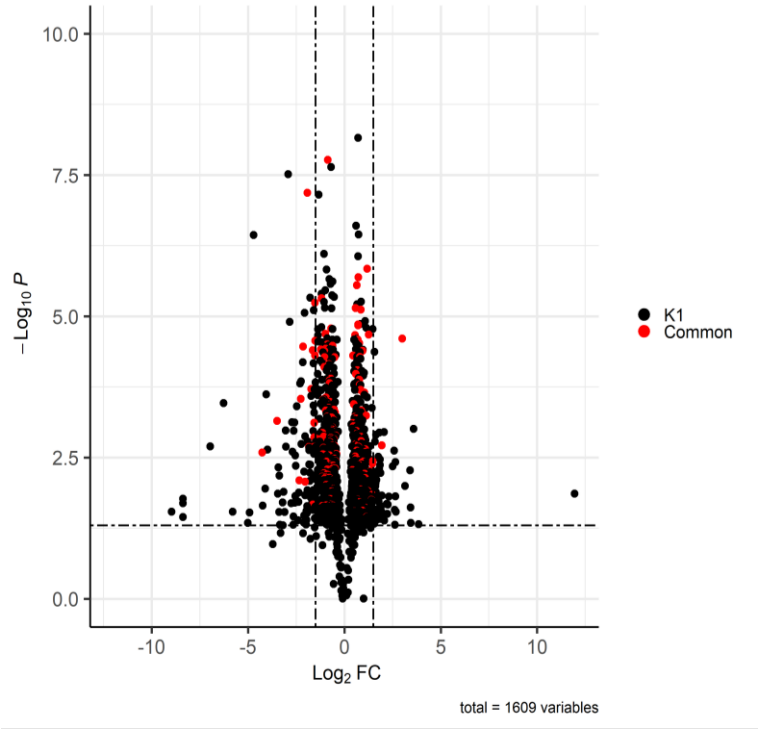


Figure 7.9: **A volcano plot of all one thousand six hundred and nine unique DE proteins detected in CHO K1 cells following adaptation from suspension growth in media containing (5%) serum to serum free (0%) media. Red dots represent the one hundred and sixty-five DE proteins common to both CHO K1 and DXB11 cells. The vertical dashed lines in Figure 7.8 represent a fold change of ± 1.5 and the horizontal dashed line represents $\log_{10}(0.05)$.**

An upset plot (Figure 7.10) was used to identify common proteins between different subpopulations of DE proteins detected in the two cells lines.

The size of each subpopulation (Set Size) is represented by the small bar chart at the bottom of Figures 7.10. The overlap between each subpopulation (Intersection Size) is represented by the y-axis of main bar chat. What each intersection is, is explained by the grid on the x-axis of the main bar chart. The number above each bar is the number of DE proteins in that set. Down.K1, and Down.DXB11 are the sets of DE proteins identified in K1 and DXB11 cells, respectively, whose expression was lower in serum free media compared to media containing serum.Up.K1, and Up.DXB11 are the sets of DE proteins identified in K1 and DXB11 cells, respectively, whose expression was lower in serum free media compared to media containing serum.

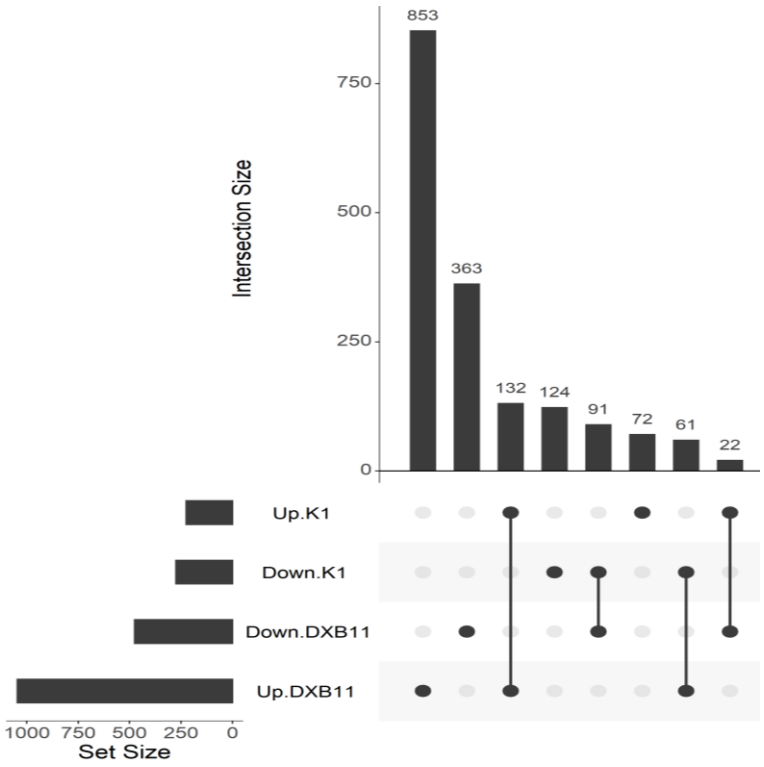


Figure 7.10: **The number of unique DE proteins present in each of the subpopulations of interest and the number of DE proteins which overlap between two or more subpopulations. The numbers above the bars refer to the number of DE proteins represented by that bar. Set Size on the lower left-hand corner of the figure refers to the total number of DE proteins in each subpopulation (e.g., there are 132 DE proteins in total in the Up.K1 subpopulation and 853 DE proteins in total in the Up.DXB11 subpopulation. However, these two subpopulations have 107 DE proteins in common.**

The subpopulations under comparison were:

a) Proteins whose abundance was higher in suspension cells adapted to growth in serum free media than suspension cells adapted to growth in media containing serum in the DXB11 cell line. These proteins were classified as having a negative correlation with the adaptation to suspension and are part of the Up.DXB11 subset.
b) Proteins whose abundance was lower in in suspension cells adapted to growth in serum free media than suspension cells adapted to growth in media containing serum in the DXB11 cell line. These proteins were classified as having a positive correlation with the adaptation to suspension and are part of the Down.DXB11 subset.

c) Proteins whose abundance was higher in suspension cells adapted to growth in serum free media than suspension cells adapted to growth in media containing serum in the K1 cell line. These proteins were classified as having a negative correlation with the adaption to suspension and are part of the Up.K1 subset.
d) Proteins whose abundance was lower in suspension cells adapted to growth in serum free media than suspension cells adapted to growth in media containing serum in the K1 cell line. These proteins were classified as having a positive correlation with the adaption to suspension and are part of the Down.K1 subset.

However, all the differentially expressed proteins presented in Figure 7.10 fell into one of eight subpopulations as described below. This makes sense as it's not possible for the same differentially expressed protein to be a member of both the upregulated DXB11 and downregulated DXB11 sub populations.

- Eight hundred and fifty three proteins whose expression was higher in suspension cells than attached cells (Up.DXB11) were found to be uniquely DE in CHO DXB11 cells.
- Three hundred and sixty three proteins whose expression was lower in suspension cells than attached cells (Down.DXB11) were found to be uniquely DE in CHO DXB11 cells.
- One hundred and twenty four proteins were present at lower levels in suspension K1 cells compared to attached cells (Down.K1). In this study they were found to be uniquely DE in CHO K1 cells.
- Seventy two proteins were up regulated in suspension K1 cells compared to attached cells (Up.K1). In this study they were found to be uniquely DE in CHO K1 cells.
- Sixty one DE proteins present in the Down.K1 subpopulation were also found to be present in the Up.DXB11 subpopulation, i.e., their expression profile differs between the two cell lines.
- Twenty two DE proteins present in the Up.K1 subpopulation were also found to be present in the Down.DXB11 subpopulation, i.e., their expression profile differs between the two cell lines.
- One hundred and thirty two DE proteins common to both CHO K1 and DXB11 cells had a higher expression level in suspension than attached.

- Ninety one DE proteins common to both CHO K1 and DXB11 cells had a lower expression level in suspension than attached.

The last two categories comprise the proteins of interest. These are the two hundred and twenty one hundred DE proteins common to K1 and DXB11 cell lines which exhibited the same expression pattern upon adaptation to growth in serum free media (Appendix 22)

These proteins were considered to be the proteins underpinning the adaptation of cells from growth in media supplemented with serum to growth in serum free media, as they displayed a consistent DE protein response in both cell lines following their adaptation to growth in serum free media.

7.3.3. Identification of common enriched KEGG pathways identified in CHO K1 and CHO DXB11 cells

PSEA, a variation of GSEA, was developed specifically for the analysis of proteomic datasets. This study used the R package PathfindR (Ulgen et al., 2019) to perform PSEA on the two hundred and twenty one DE proteins identified as being common to both K1 and DXB11 cells under the conditions being evaluated. The analysis was performed separately on each cell line and the results combined. This approach was taken because although the proteins are common to both K1 and DXB11 cell lines the fold change and p value for each protein differed between the two cells lines. The output of the analysis was a list of statistically enriched KEGG pathways, the DE proteins detected in those paths and whether these proteins were upregulated or downregulated. The output from each pathfindR analysis were compared to see if the enriched KEGG path IDs differed between K1 and DXB11 cell lines and only pathways present in both cell lines were considered further.

This resulted in the identification of sixty-one KEGG pathways as being enriched. Nine were unique to DXB11, fifteen were unique to K1 and the remaining thirty-seven were common to both cell lines (Appendix 23). Fifty two of the two hundred and twenty one DE proteins common to both CHO K1 and DXB11 cells lines were found to be present in these enriched pathways. Sixteen of these DE proteins were found to be downregulated while thirty six were found to be upregulated.

The KEGG system is collection of pathway maps sub divided into seven common themes. Therefore, by identifying the individual KEGG pathways which are statistically enriched in the comparison an overall picture of what changed within the cell during the adaptation can be obtained. Figure 7.11, shows the distribution of the sixty-two significantly enriched KEGG pathways detected following the adaptation of K1 and DXB11 cell lines from growth in media containing serum to serum free media across the seven themes which comprise the KEGG system.

As a theme, human diseases (17/62 pathways) dominated the list of enriched KEGG pathways identified following the adaptation of K1 and DXB11 cells from attached to suspension growth. While the link between diseases such as cancer and adhesion is well established it's difficult to see a link between human disease as a category and the adaption of mammalian cell lines to growth in serum free media. Hence this category was not considered in any further analysis. Organismal systems (Figure 7.12) and metabolism (Figure 7.13) ranked joint second (14/62).

The organismal systems of interest detected in this dataset included aging, digestion, and the endocrine system. Pathways and proteins detected within these systems included, Longevity regulating pathway (HAPA8), carbohydrate digestion (HK1) and estrogen signalling (HSP90A1, FKBP4, HAPA8 and ITPR3). Cellular processes identified as being enriched included regulation of the actin cytoskeleton, tight and gap junctions, as well as processes involved in catabolism such as phagosome and endocytosis.

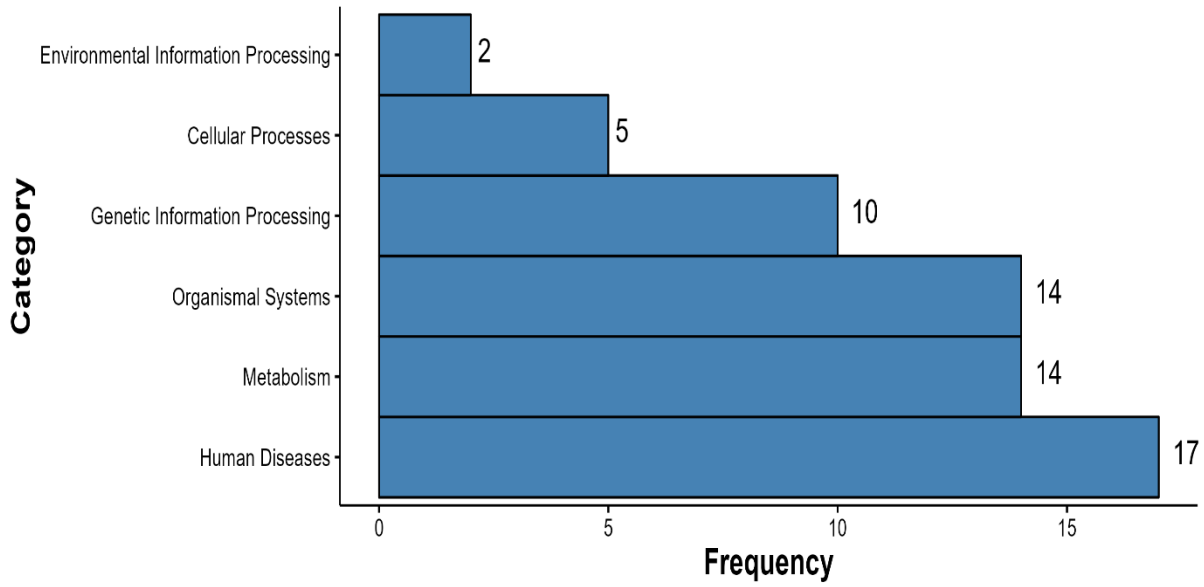


Figure 7.11: **Categorisation of significantly enriched KEGG pathways detected following the adaption of CHO K1 and DXB11 cell lines from suspension growth media containing serum to suspension growth in serum free media. The enriched KEGG pathways have been grouped by KEGG category. Frequency refers to the number of enriched KEGG pathways detected within each category.**

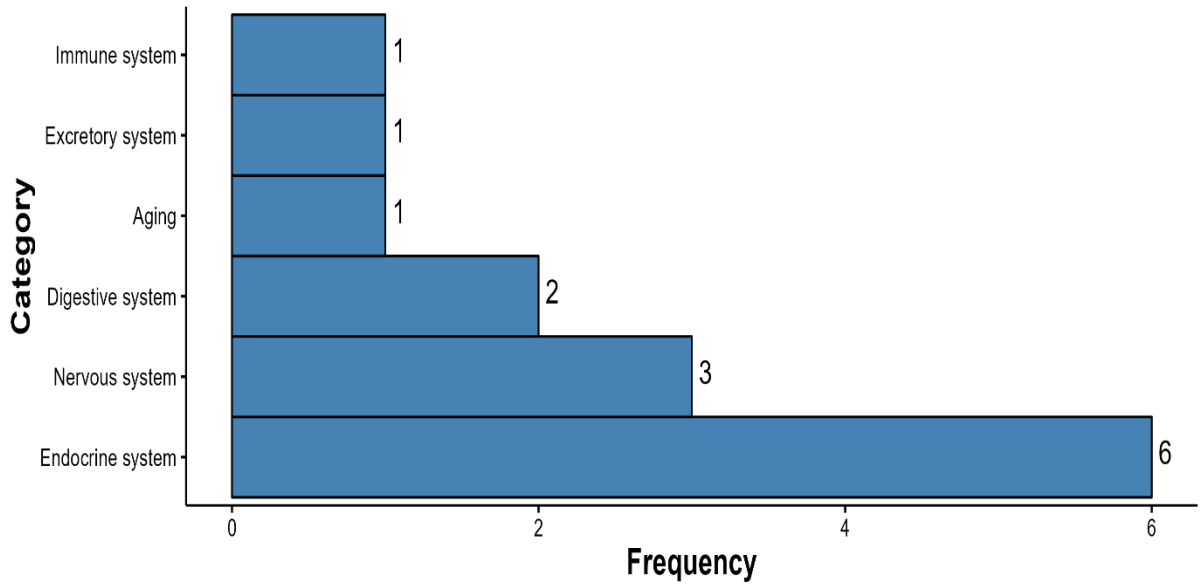


Figure 7.12: **All significantly enriched KEGG pathways detected following the adaption of CHO K1 and DXB11 cell lines from suspension growth in media containing serum to suspension growth in serum free media which belong to the organismal systems category from Figure 7.11. Frequency refers to the number of enriched KEGG pathways detected within each category.**

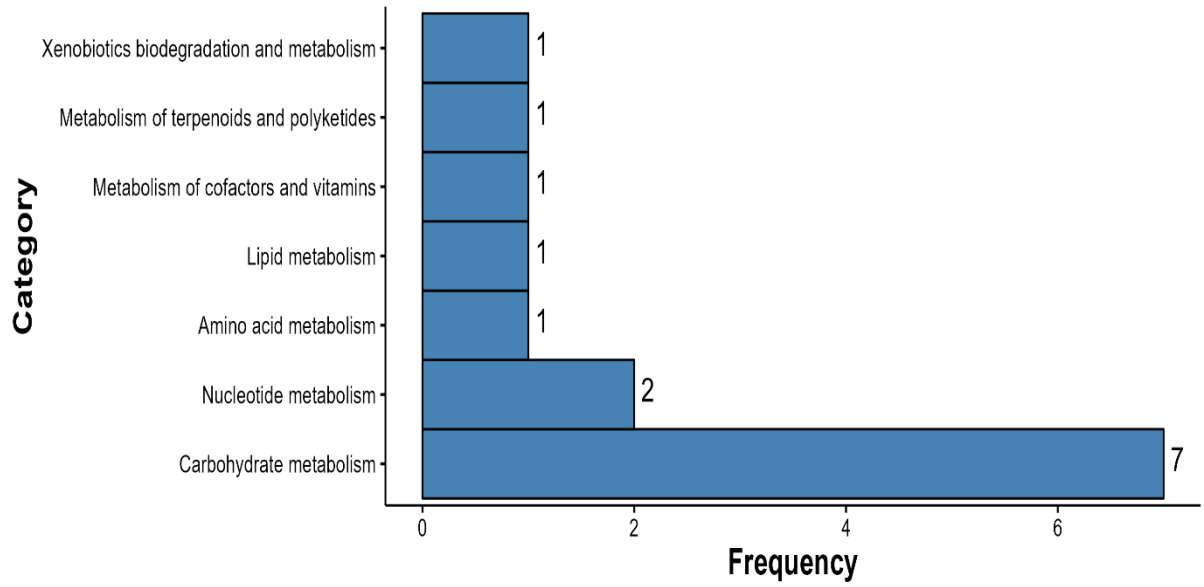


Figure 7.13: **All significantly enriched KEGG pathways detected following the adaption of CHO K1 and DXB11 cell lines from suspension growth in media containing serum to suspension growth in serum free media which belong to the metabolism category from Figure 7.11. Frequency refers to the number of enriched KEGG pathways detected within each category.**

As a category metabolism can be broken down into various subgroups. Metabolism of carbohydrates was the largest single category (7/14 pathways) while the metabolism of alternative energy sources (lipids and amino acids) accounted for (2/14 pathways).

The ten most enriched pathways ordered by significance (p value) and fold enrichment based on the results of the PSEA analysis of the one hundred and sixty-five DE proteins identified in this study are displayed in Figure 7.14. Figure 7.14 combines information about the number of proteins detected in each pathway (bubble size) with the degree of enrichment (x-axis) and how significant that pathway is (bubble colour). What it doesn't do is capture whether the proteins identified as being part of those pathways were upregulated or downregulated or indeed what those proteins are. That information is presented in Table 7.5. An additional criterion was applied in the generation of this Table 7.5. PSEA analysis was performed on the DE proteins from K1 and DXB11 cells separately. This was because although the proteins were common to both cell lines the p value associated with them differed as the two cell lines were analysed in separate MS experiments. Therefore, only KEGG pathways identified as being enriched in both cell lines were reported in this table.

Fifty-four of the one hundred and sixty-five common DE proteins were identified as being components of the KEGG paths displayed in Figure 7.14 and Table 7.5. Seventeen of these proteins were downregulated upon the removal of serum from the growth media, while the other thirty-seven were upregulated.

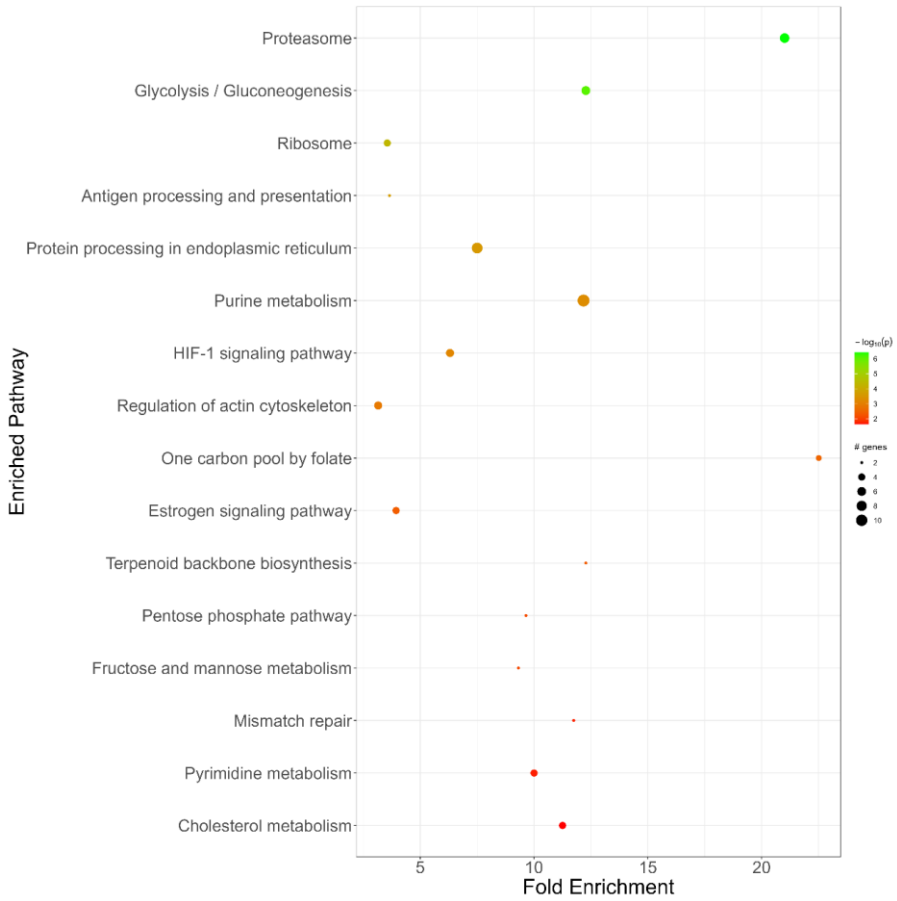


Figure 7.14: **Identification of the top 10 enriched pathways ranked by fold change and top 10 ranked by significance common to both CHO K1 and CHO DXB11 following adaptation to growth in serum free media. The colour of the bubble represents the p value associated with that pathway, while the size of the bubble represents the number of DE proteins identified in that pathway based on analysis by pathfindR. Some pathways are common to both categories hence why only sixteen are plotted.**

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated	Down Regulated
Top 10 KEGG Paths ordered by significance					
hsa03050	Proteasome	21.01	$2.43E10^{-18}$	PSMD6, PSMC1, PSMC6, PSMC3, PSMA6, PSMB3, PSMB5	NA
hsa00010	Glycolysis / Gluconeogenesis	12.28	$8.39E10^{-07}$	HK1, ALDOA, PGK1, ENO1, PKM	ALDH2
hsa04066	HIF-1 signalling pathway	6.31	$5.98E10^{-06}$	HMOX1, HK1, ALDOA, ENO1,	NA
hsa03430	Mismatch repair	11.75	$6.43E10^{-06}$	PCNA, RPA1	NA
hsa04141	Protein processing in endoplasmic reticulum	7.50	$1.90E10^{-05}$	HSPA8, HSP90AB1, RAD23B	CKAP4, RRBP1, SIL1, DNAJC10, EDEM3, PDIA6
hsa03010	Ribosome	3.55	$2.46E10^{-05}$	RPL4, RPL5, RPL18	MRPS9
hsa00240	Pyrimidine metabolism	10.01	$1.57E10^{-4}$	NME2, DUT, PNP	RRM1
hsa04612	Antigen processing and presentation	3.65	$2.34E10^{-04}$	HSPA8, HSP90AB1	NA
hsa00230	Purine metabolism	12.18	$5.10E10^{-04}$	PRPS1, GART, PFAS, PAICS, ATIC, PNP, HPRT1, NME2, PKM, ADK	RRM1

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated	Down Regulated
hsa04810	Regulation of actin cytoskeleton	3.16	$1.32E10^{-04}$	RDX, CFL1, IQGAP1	IQGAP2, GSN
Top 10 KEGG paths ordered by fold enrichment					
hsa00670	One carbon pool by folate	22.51	3.02E-03	MTHFD1, GART, ATIC	NA
hsa03050	Proteasome	21.01	3.46E-16	PSMD6, PSMC1, PSMC6, PSMC3, PSMA6, PSMB3, PSMB5	NA
hsa00010	Glycolysis / Gluconeogenesis	12.28	8.39E-07	HK1, ALDOA, PGK1, ENO1, PKM	ALDH2
hsa00900	Terpenoid backbone biosynthesis	12.28	4.56E-03	FDPS	HMGCS2
hsa00230	Purine metabolism	12.18	5.10E-04	PRPS1, GART, PFAS, PAICS, ATIC,	RRM1
hsa03430	Mismatch repair	11.75	4.99E-03	PCNA, RPA1	NA
hsa04979	Cholesterol metabolism	11.26	2.32E-02	NA	NCEH1, SCARB1, LRP1, VAPA
hsa00240	Pyrimidine metabolism	10.01	1.57E-04	NME2, DUT, PNP	NA
hsa00030	Pentose phosphate pathway	9.65	7.46E-03	PRPS1, ALDOA	NA

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated	Down Regulated
hsa00051	Fructose and mannose metabolism	9.32	8.01E-03	HK1, ALDOA	NA
For space reasons HGNC symbols rather than protein names have been used throughout this table. Protein names are listed in Table 7.14 at the end of this chapter.					

Table 7.5: **The ten KEGG pathways displaying the greatest degree of enrichment or significance following PSEA analysis of CHO K1 and CHO DXB11 cells adapted from growth in media containing serum to growth in serum free media.**

Pathways related to metabolism dominate Figure 7.14 and Table 7.5, account for 7/15 pathways. The enriched metabolic pathways detected following the adaptation of CHO cells from suspension growth in media containing serum to serum free media were related to *carbohydrate metabolism* (3/15)*, nucleotide metabolism* (2/15) and *cholesterol synthesis or metabolism* (2/15). Pathways related to cell signalling, cell motility and protein folding were also detected as being among the ten most enriched pathways based on either statistical significance or fold change detected in the data.

Five of the pathways presented in Table 7.5 were present among both the ten most significant pathways as well as the ten pathways with the greatest fold change. These were *proteasome, glycolysis/glucogenesis*, *purine and pyrimidine metabolism* and *mismatch repair*.

Additionally, in the context of the overall theme of this thesis it's worth noting that *regulation of the actin cytoskeleton,* a key adhesion pathway, was among the ten most significantly enriched pathways.

One of the aims behind performing a proteomic analysis of CHO K1 and CHO DXB11 cells following their adaptation from growth in media containing serum to growth in serum free media was to try and gain an insight into what changes in protein expression correlated with this adaptation.

7.3.4. Gene Ontology analysis of DE proteins common to CHO K1 and DXB11 cells

Gene ontology (GO) describes the biological processes (BP), molecular function (MF) and cellular component (CC) associated with an individual gene. Given the link between genes and proteins it can also be used to analyse protein lists. It's designed to be species agnostic and therefore a *Homo sapiens* reference was used as the background dataset for this analysis. The R packages BiomaRt (Durinck et al., (2009) and Annotation Hub (Morgan and Shepard, 2021) were used to assigned gene ontologies to the eighty-one proteins identified by PSEA and identify which of GO terms were statistically enriched.

Gene ontology analysis of the eighty-one differentially expressed proteins present in the enriched KEGG pathways revealed an association with five hundred and nineteen different GO terms (Appendix 24). Table 7.6 lists the five most significantly enriched GO

terms in each ontology associated with the enriched KEGG pathways identified by the PSEA analysis (Table 7.5).

ID	Description	p value	p adjusted	q value	# Proteins	Ontology
GO:0006163	purine nucleotide metabolic process	6.00E-12	1.00E-08	6.79E-09	16	BP
GO:0072521	purine-containing compound metabolic process	1.13E-11	1.00E-08	6.79E-09	16	BP
GO:0009165	nucleotide biosynthetic process	1.66E-11	1.00E-08	6.79E-09	13	BP
GO:1901293	nucleoside phosphate biosynthetic process	1.91E-11	1.00E-08	6.79E-09	13	BP
GO:0043312	neutrophil degranulation	2.49E-11	1.00E-08	6.79E-09	16	BP
GO:0045296	cadherin binding	4.02E-10	1.17E-07	7.87E-08	13	MF
GO:0005095	GTPase inhibitor activity	1.55E-05	0.001758	0.001187	3	MF
GO:0033218	amide binding	1.93E-05	0.001758	0.001187	9	MF
GO:0042277	peptide binding	2.88E-05	0.001758	0.001187	8	MF
GO:0023026	MHC class II protein complex binding	3.64E-05	0.001758	0.001187	3	MF
GO:1904813	ficolin-1-rich granule lumen	2.47E-11	5.89E-09	3.92E-09	10	CC
GO:0101002	ficolin-1-rich granule	1.28E-09	1.53E-07	1.02E-07	10	CC
GO:0034774	secretory granule lumen	2.46E-07	1.39E-05	9.27E-06	10	CC
GO:0060205	cytoplasmic vesicle lumen	2.76E-07	1.39E-05	9.27E-06	10	CC
GO:0031983	vesicle lumen	2.92E-07	1.39E-05	9.27E-06	10	CC

Table 7.6: **Significantly enriched GO terms following PSEA analysis of CHO K1 and CHO DXB11 cells adapted from growth in media containing serum to growth in serum free media. Adjusted p value and q Value are the p value following application of the Benjamini-Hochberg formula and false discovery rate (FDR) respectively to the calculated p value.**

As can been seen from Table 7.6, biological processes dominate the list in terms of significance and suggest that transcription and translation via the synthesis and metabolism of nucleotide and nucleosides played a significant role or were particularly perturbed during the adaption of CHO cells from suspension growth in media containing serum to serum free media. The most significantly enriched molecular functions included cadherin binding which supports the finding that regulation of the actin cytoskeleton was a significantly enriched KEGG pathway. The list of most significantly enriched cellular components was dominated by ficolin granules. Ficolin's are part of the innate immune system and while this may suggest that the process of wound healing is a good model for describing the events involved during the adaptation of CHO cells to growth in serum free media the data presented in Table 7.8 doesn't support this. Secretory granules of the lumen also featured on the list of significantly enriched cellular components. This finding along with the observation that peptide and amide binding were significantly enriched molecular functions supports the idea that protein catabolism maybe involved in the adaptation of CHO cells to growth in serum free media as this occurs in the lumen of cells. Whether this catabolism is due to cells dying as they fail to adapt to changes in the culture media (i.e., the withdrawal of serum) or is providing amino acids as an alternative energy source or the process of synthesising essential amino acids is not clear.

7.3.5. Identification of DE proteins and KEGG pathways common CHO K1 and DXB11 cells which are related to cellular adhesion

However, in keeping with the overall aim of this thesis, i.e., the development of a flocculating CHO cell line, and the approach taken in chapter six it was decided to search the eighty-one DE proteins identified by the PSEA analysis against the EMT (Appendix 14) and wound healing (Appendix 15) models to try and identify additional DE proteins involved in adhesion to those identified in chapter six. The results of these filtering steps are presented in Table 7.7 and Table 7.8 respectively. These represent previously published proteins associated with EMT and wound healing which were among the DE proteins identified as being present in the sixty-two enriched KEGG pathways common to both CHO K1 and CHO DXB11 cells adapted from growth in media containing serum to serum free media.

HGNC Symbol	Protein Name	Uniprot ID	Status	FC K1	FC DXB11
ADAM9	ADAM metallopeptidase domain 9	Q13443	Down	0.515	0.641
ANLN	anillin actin binding protein	Q9NQW6	Up	1.821	2.721
APMAP	adipocyte plasma membrane associated protein	Q9HDC9	Down	0.586	0.716
ARHGEF40	Rho guanine nucleotide exchange factor 40	Q8TER5	Down	0.645	0.286
CBX5	chromobox 5	P45973	Up	1.669	1.793
CCT2	chaperonin containing TCP1 subunit 2	P78371	Up	1.603	1.997
CCT3	chaperonin containing TCP1 subunit 3	P49368	Up	1.568	1.941
CDV3	CDV3 homolog	Q9UKY7	Up	2.261	2.42
CLIC1	chloride intracellular channel 1	O00299	Up	2.056	2.041
COL5A1	collagen type V alpha 1 chain	P20908	Down	0.122	0.179
COL6A1	collagen type VI alpha 1 chain	P12109	Down	0.132	0.167
DCXR	dicarbonyl and L-xylulose reductase	Q7Z4W1	Down	0.583	0.224
DHPS	deoxyhypusine synthase	P49366	Up	1.739	7.245
DUT	deoxyuridine triphosphatase	P33316	Up	1.759	1.608
EHD2	EH domain containing 2	Q9NZN4	Down	0.596	0.458
EIF1AX	eukaryotic translation initiation factor 1A X-linked	P47813	Up	1.908	3.555
ERAP1	endoplasmic reticulum aminopeptidase 1	Q9NZ08	Down	0.424	0.6
FAAH	fatty acid amide hydrolase	O00519	Down	0.493	0.275

HGNC Symbol	Protein Name	Uniprot ID	Status	FC K1	FC DXB11
FDPS	farnesyl diphosphate synthase	P14324	Up	1.576	3.115
GALNT2	polypeptide N-acetylgalactosaminyltransferase 2	Q10471	Down	0.524	0.598
GPD2	glycerol-3-phosphate dehydrogenase 2	P43304	Up	1.632	2.935
GSN	Gelsolin	P06396	Down	0.513	0.329
HMOX1	heme oxygenase 1	P09601	Up	2.194	2.963
HSD17B11	hydroxysteroid 17-beta dehydrogenase 11	Q8NBQ5	Down	0.593	0.311
ICAM1	intercellular adhesion molecule 1	P05362	Up	1.646	1.383
IKBIP	IKBKB interacting protein	Q70UQ0	Down	0.563	0.33
KPNA2	karyopherin subunit alpha 2	P52292	Up	2.239	1.625
LIMS1	LIM zinc finger domain containing 1	P48059	Up	1.907	1.912
LRP1	LDL receptor related protein 1	Q07954	Down	0.573	0.216
MAOA	monoamine oxidase A	P21397	Down	0.265	0.314
MAP1B	microtubule associated protein 1B	P46821	Up	1.842	2.027
MICAL1	microtubule associated monooxygenase, calponin and LIM domain containing 1	Q8TDZ2	Down	0.627	0.408
MTA1	metastasis associated 1	Q13330	Down	0.501	0.749
MTHFD1	methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1	P11586	Up	1.564	2.073
MYOF	Myoferlin	Q9NZM1	Down	0.542	0.813

HGNC Symbol	Protein Name	Uniprot ID	Status	FC K1	FC DXB11
NCEH1	neutral cholesterol ester hydrolase 1	Q6PIU2	Down	0.663	0.403
NQO1	NAD(P)H quinone dehydrogenase 1	P15559	Down	0.572	0.13
PARK7	Parkinsonism associated deglycase	Q99497	Up	2.392	3.451
PCNA	proliferating cell nuclear antigen	P12004	Up	1.653	1.436
PDIA6	protein disulfide isomerase family A member 6	Q15084	Down	0.633	0.551
PHLDB1	pleckstrin homology like domain family B member 1	Q86UU1	Down	0.488	0.429
PKM	pyruvate kinase M1/2	P14618	Up	1.622	1.539
PSAT1	phosphoserine aminotransferase 1	Q9Y617	Up	1.55	1.944
PSMC6	proteasome 26S subunit, ATPase 6	P62333	Up	3.59	1.977
PTGR1	prostaglandin reductase 1	Q14914	Up	1.635	2.379
RAB31	RAB31, member RAS oncogene family	Q13636	Down	0.525	0.463
RPL5	ribosomal protein L5	P46777	Up	2.046	5.137
RRBP1	ribosome binding protein 1	Q9P2E9	Down	0.655	0.56
SNX4	sorting nexin 4	O95219	Down	0.562	0.203
SPARC	secreted protein acidic and cysteine rich	P09486	Down	0.486	0.007
SRM	spermidine synthase	P19623	Up	2.713	2.141
SSRP1	structure specific recognition protein 1	Q08945	Up	1.761	1.551
TGM2	transglutaminase 2	P21980	Down	0.559	0.159

HGNC Symbol	Protein Name	Uniprot ID	Status	FC K1	FC DXB11
TJP2	tight junction protein 2	Q9UDY2	Up	1.647	2.505
TNS1	tensin 1	Q9HBL0	Down	0.346	0.224
TXNRD1	thioredoxin reductase 1	Q16881	Up	1.67	2.491
VAPA	VAMP associated protein A	Q9P0L0	Down	0.601	0.491

Table 7.7: DE proteins identified as being common to both K1 and DXB11 cells after adaptation from growth in media containing serum to growth in serum free media which are also known to be involved in EMT.

Fifty-eight proteins identified as being common to both K1 and DXB11 cells in this study had been previously published by other groups as being associated with the EMT. This represented over a third of the one hundred and sixty-five DE proteins identified in this study as being common to K1 and DXB11 cells after adaptation to growth in serum free media. However, it's almost three quarters of the 81 DE proteins identified as being present in the enriched KEGG pathways identified in this study as being common to K1 and DXB11 cells after adaptation to growth in serum free media.

HGNC Symbol	Protein Name	Uniprot.ID	Status	FC K1	FC DXB11
ADAM9	ADAM metallopeptidase domain 9	Q13443	Down	0.515	0.641
EIF3E	eukaryotic translation initiation factor 3 subunit E	P60228	Up	1.756	1.776
HMOX1	heme oxygenase 1	P09601	Up	2.194	2.963
ICAM1	intercellular adhesion molecule 1	P05362	Up	1.646	1.383
SPARC	secreted protein acidic and cysteine rich	P09486	Down	0.486	0.007

Table 7.8: DE proteins identified as being common to both K1 and DXB11 cells after adaptation from growth in media containing serum to growth in serum free media which are also known to be involved in wound healing.

In comparison to the EMT data (Table 7.7) only five proteins identified in this study as being common to both K1 and DXB11 cells following adaptation growth in serum free media been previously published by other groups as have an association with the biological process of wound healing (Table 7.8). The statement that the proteins listed in Tables 7.7 and 7.8 have been previously published as having an association with either the EMT or process of wound healing is based on their presence in the EMTome (Vasaikar et al., 2021) and TiRe (Yanai et al., 2016) databases. These two databases contain eighty-four and three hundred and eleven publications respectively, describing various proteins implicated in either the EMT or wound healing.

Four proteins were common to both Tables 7.7 and 7.8. These were *ADAM metallopeptidase domain 9 (ADAM 9), intercellular adhesion molecule 1 (ICAM1), heme oxygenase 1 (HMOX1)* and *secreted protein acidic and cysteine rich (SPARC)*. ADAM9 is a transmembrane protein known to regulate cellular adhesion via its interaction with integrins. ICAM1 (aka CD54) is a member of the Immunoglobulin super family of adhesion proteins which is known to mediate the adhesion of leukocytes and endothelial cells. HMOX1 is known to modulate the expression of endothelial cell

adhesion molecules such as ICAM and protects endothelial cells from undergoing apoptosis while SPARC is an extracellular matrix protein which mediates the disruption of focal adhesions.

To gain further insight into the proteins/KEGG pathways involved in cellular adhesion, the entire KEGG database was searched. The results of this search were used to build an R vector containing forty-nine individual entries (Appendix 16) which was then used to filter the list of sixty-one statistically enriched KEGG pathways listed in Appendix 23. The results of this filtering step are presented in Table 7.9 and represent the adhesion related KEGG pathways which were statistically enriched during the adaption of CHO K1 and CHO DXB11 cells to growth in serum free media.

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated DE Proteins	Down Regulated DE Proteins
hsa04810	Regulation of actin cytoskeleton	3.14	0.001	RDX, CFL1, IQGAP1	IQGAP2, GSN
hsa04915	Estrogen signalling pathway	2.91	0.016	HSP90AB1, FKBP4, HSPA8	ITPR3
Protein names are listed in Table 7.14 at the end of this chapter.					

Table 7.9: **All adhesion related KEGG pathways found to be significantly enriched following adaptation of CHO K1 and CHO DXB11 cells from growth in media containing serum to growth in serum free media. For space reasons HGNC symbols rather than protein names have been used.**

Only two KEGG pathways (Table 7.9) associated with adhesion were found to be statistically enriched during adaption of CHO K1 and DXB11 cells from growth in media containing serum to serum free media. This suggests that despite the data presented in Table 7.7 and Table 7.8 adhesion is not the primary biological process being disrupted during the adaptation of CHO K1 and DXB11 cells to growth in serum free media.

7.3.6. Impact of correcting for false discovery rates (FDR) on the identification of differentially expressed proteins common to CHO K1 and CHO DXB11 cells after adaptation to growth in serum free media

As discussed in chapter six, during the review of this thesis prior to submission the lack of an FDR criteria during selection of DE proteins during the Progenesis QI for Proteomics workflow was commented upon. Accordingly, the DE protein list generated by Progenesis QI for Proteomics has been interrogated following application of a retrospective FDR criteria at this step and an assessment of the impact omitting such a criteria had in the identification of DE proteins has been made. This assessment is discussed in the following paragraphs.

One thousand one hundred and eighty DE proteins were detected in CHO DXB11 cells compared to one thousand six hundred and nine in CHO K1 cells following adaptation to suspension growth in serum free media.

Following application of a false discovery rate (q value) of 0.01 at the DE peptide selection step in Progenesis Qi for Proteomics, the number of DE proteins identified in CHO K1 cells was reduced to three hundred and forty-eight (Appendix 25) and to one hundred and twenty-one in DXB11 cells (Appendix 26). An Upset plot identified thirty DE proteins common to both CHO K1 and CHO DXB11 cells. These thirty DE proteins were analysed by peptide set enrichment analysis (PSEA) in order to identify the enriched KEGG pathways they were implicated in. Enriched gene ontology terms associated with these thirty DE proteins were identified using the Annotation Hub package in R.

Table 7.10 lists the top five proteins in either cell line based on:

c. Absolute fold change (absolute FC) between the two experimental groups
d. p value

However, it was decided to focus on the thirty five common DE proteins identified using an Upset plot (Figure 7.15), which are either upregulated (Up.DXB11 and Up.K1) or downregulated (Down.DXB11 and Down.K1) in both CHO K1 and CHO DXB11 cells. These thirty five proteins are more applicable to explaining the events underpinning the adaptation of CHO K1 and CHO DXB11 cells to growth in serum free media.

Cell Line	Protein Name	HGNC Symbol	p Value	Max FC[1]	Log_2 (Absolute FC)[2]	Present in all attached reps	Present in all suspension reps
Ranked by absolute FC							
K1	small proline rich protein 1A	SPRR1A	$2.50E10^{-5}$	8.03	3.01	Yes	Yes
	eukaryotic translation elongation factor 1 gamma	EEF1G	$2.39E10^{-3}$	6.02	2.59	Yes	Yes
	DLG associated protein 5	DLGAP5	$5.91E10^{-3}$	3.68	1.88	Not detected in S0.5c_early, S0.5d_early	Yes
	proteasome 26S subunit, ATPase 6	PSMC6	$3.43E10^{-3}$	3.59	1.84	Yes	Yes
	Histone H2B type 1 (Sumatran Orangutan)	H2B1-PONAB	$2.86E10^{-2}$	3.31	1.73	Yes	Yes
DXB11	Serine/threonine protein phosphatase 2 scaffold	PPP2R1B	$2.48E10^{-5}$	12.1	2.38	Yes	Yes
	PDZ and LIM domain 1	PDLIM1	$5.22E10^{-5}$	6.60	2.29	Yes	Yes
	actin gamma 1	ACTG1	$1.53E10^{-2}$	6.98	2.23	Yes	Yes
	Asparagine--tRNA ligase, cytoplasmic	NARS	$2.39E10^{-3}$	7.46	2.20	Yes	Yes
	acetyl-CoA acetyltransferase 2	ACAT2	$9.36E10^{-3}$	8.09	1.97	Yes	Yes
Ranked by p Value							
K1	annexin A1	ANXA1	$6.95E10^{-9}$	2.18	1.12	Yes	Yes
	prostaglandin reductase 1	PTGR1	$1.70E10^{-8}$	1.64	0.71	Yes	Yes
	ERGIC and golgi 3	ERGIC3	$3.06E10^{-8}$	1.61	-0.69	Yes	No
	collagen type VI alpha 1 chain	COL6A1	$6.51E10^{-8}$	7.56	-2.92	Yes	Yes
	filamin A	FLNA	$3.56E10^{-7}$	1.52	0.60	Yes	Yes
DXB11	LIM and SH3 protein 1	LASP1	$6.28E10^{-9}$	3.31	-1.46	Yes	Yes

DXB11	von Willebrand factor A domain containing 5A	VWA5A	$1.19E10^{-8}$	28.90	-4.34	Not detected in S5b, S1c, S0.5d_late	Not detected in S0a, S0b, S0c
	AHNAK nucleoprotein	AHNAK	$1.56E10^{-8}$	3.00	-1.28	Yes	Yes
	DExH-box helicase 30	DHX30	$6.67E10^{-8}$	3.09	1.37	Yes	Yes
	Plectin	PLEC	$3.50E10^{-7}$	2.40	-0.97	Yes	Yes

[1.] Max FC is the proteins fold change value as calculated by Progenesis QI for Proteomics.

[2.] Log2 Absolute FC allows a direction to be placed on the fold change. A value > 0 indicated the protein is upregulated in the comparison while a value < 0 indicates the protein is downregulated. It was calculated by dividing the proteins mean abundance in suspension cells by its mean abundance in attached cells.

Table 7.10: **The five most significant DE proteins detected in CHO K1 and CHO DXB11 cell lines as ranked by max fold change and statistical significance following application of a q value ≤ 0.01 at the DE protein level in the Progenesis QI for Proteomics workflow.**

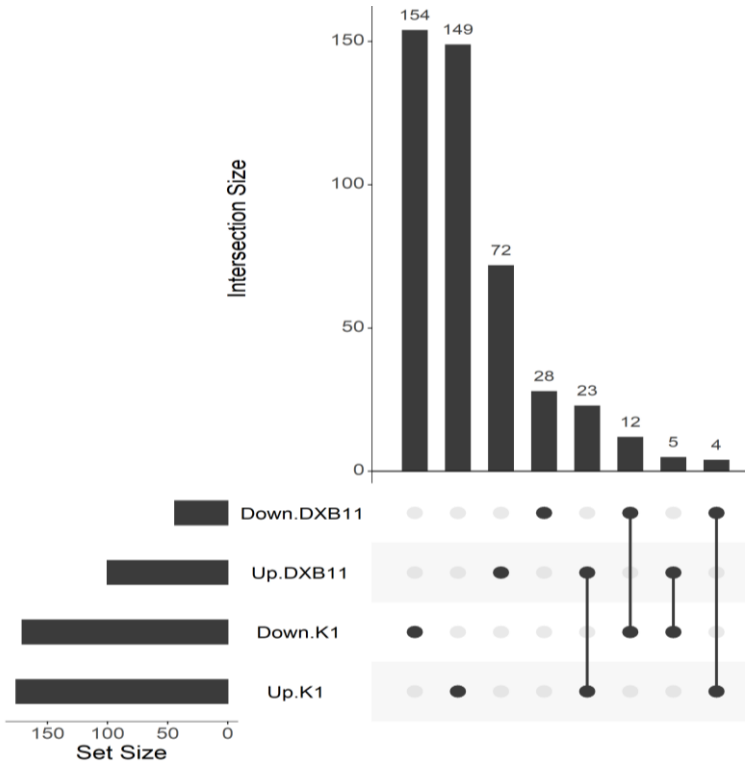


Figure 7.15: **Number of unique DE proteins present in each of the subpopulations of interest and the number of DE proteins which overlap between two or more subpopulations. The numbers above the bars refer to the number of DE proteins represented by that bar. Set Size on the lower left-hand corner of the figure refers to the total number of DE proteins in each subpopulation (e.g., there are 154 DE proteins in total in the Down.K1 subpopulation and 28 DE proteins in total in the Down.DXB11 subpopulation). However, these two subpopulations have 12 DE proteins in common.**

PSEA analysis of these thirty five DE proteins identified twenty-nine KEGG pathways as being enriched. Twenty DE proteins were identified as being a part of these pathways, nineteen of which are listed in Table 7.9.

Figures 7.16 and 7.17 show the distribution of the twenty-nine statistically enriched KEGG pathways detected following the adaptation of K1 and DXB11 cell lines from growth in media containing serum to serum free media across the seven themes which comprise the KEGG system and within the metabolic theme respectively.

As a theme, human diseases (13/32 pathways) dominated the list of enriched KEGG pathways identified following the adaptation of K1 and DXB11 cells from attached to suspension growth. An example of an enriched KEGG pathway which falls into this category was fluid shear stress and atherosclerosis. Metabolism ranked second (7/32 pathways) while organismal systems ranked third (6/32).

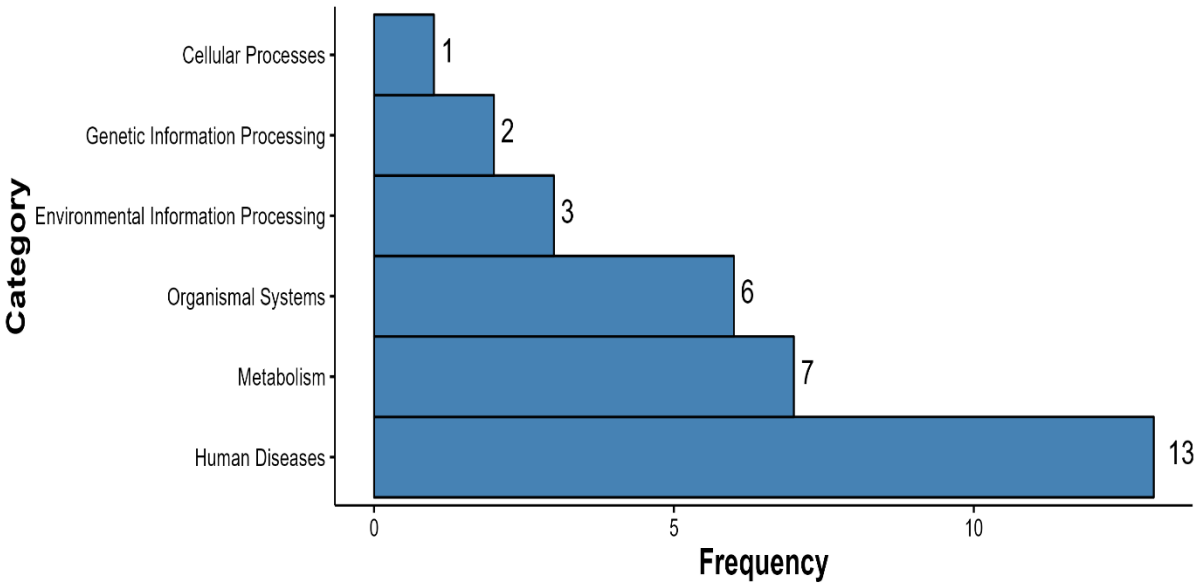


Figure 7.16: **Categorisation of significantly enriched KEGG pathways detected following the adaption of CHO K1 and DXB11 cell lines from suspension growth media containing serum to suspension growth in serum free media. The enriched KEGG pathways have been grouped by KEGG category. Frequency refers to the number of enriched KEGG pathways detected within each category.**

As a category metabolism can be broken down into various subgroups. Metabolism of amino acids was the largest single category (2/7 pathways) while the metabolism of energy, lipids and carbohydrates combined accounted for (3/7 pathways).

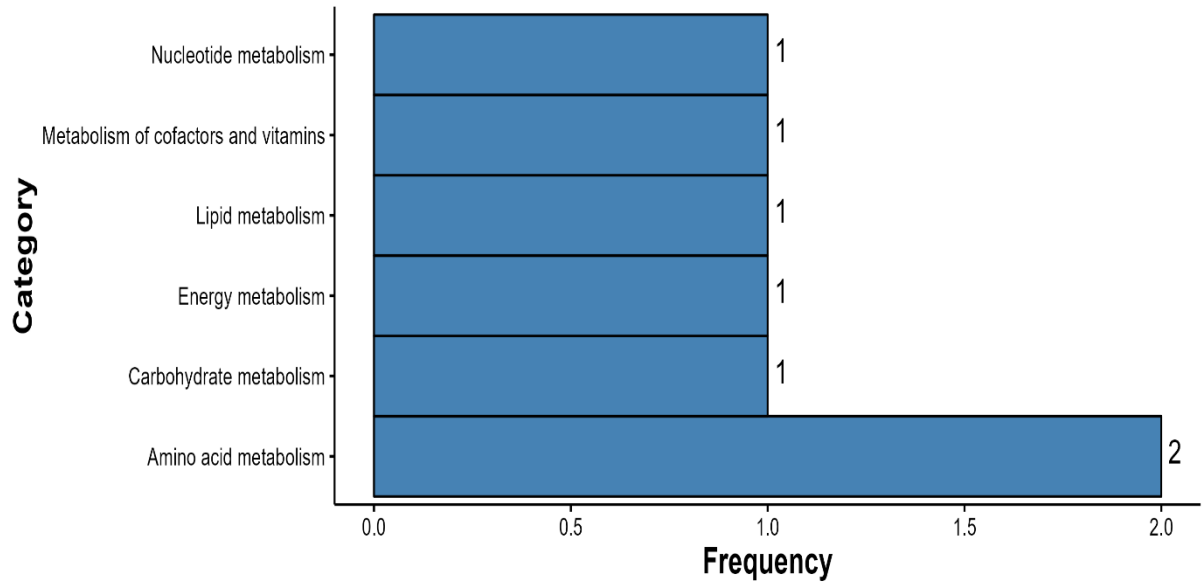


Figure 7.17: **All significantly enriched KEGG pathways detected following the adaption of CHO K1 and DXB11 cell lines from suspension growth in media containing serum to suspension growth in serum free media which belong to the metabolism category from Figure 7.18. Frequency refers to the number of enriched KEGG pathways detected within each category.**

The ten most enriched pathways ordered by significance (p value) and fold enrichment based on the results of the PSEA analysis of the one hundred and sixty-five DE proteins identified in this study are displayed in Figure 7.16. Figure 7.16 combines information about the number of proteins detected in each pathway (bubble size) with the degree of enrichment (x-axis) and how significant that pathway is (bubble colour). What it doesn't do is capture whether the proteins identified as being part of those pathways were upregulated or downregulated or indeed what those proteins are. That information is presented in Table 7.9.

All twenty common DE proteins were identified as being components of the KEGG paths displayed in Figure 7.18. Nine of these proteins were positively correlated (downregulated) with the removal of serum, while the other eleven were negatively correlated (upregulated) with the removal of serum.

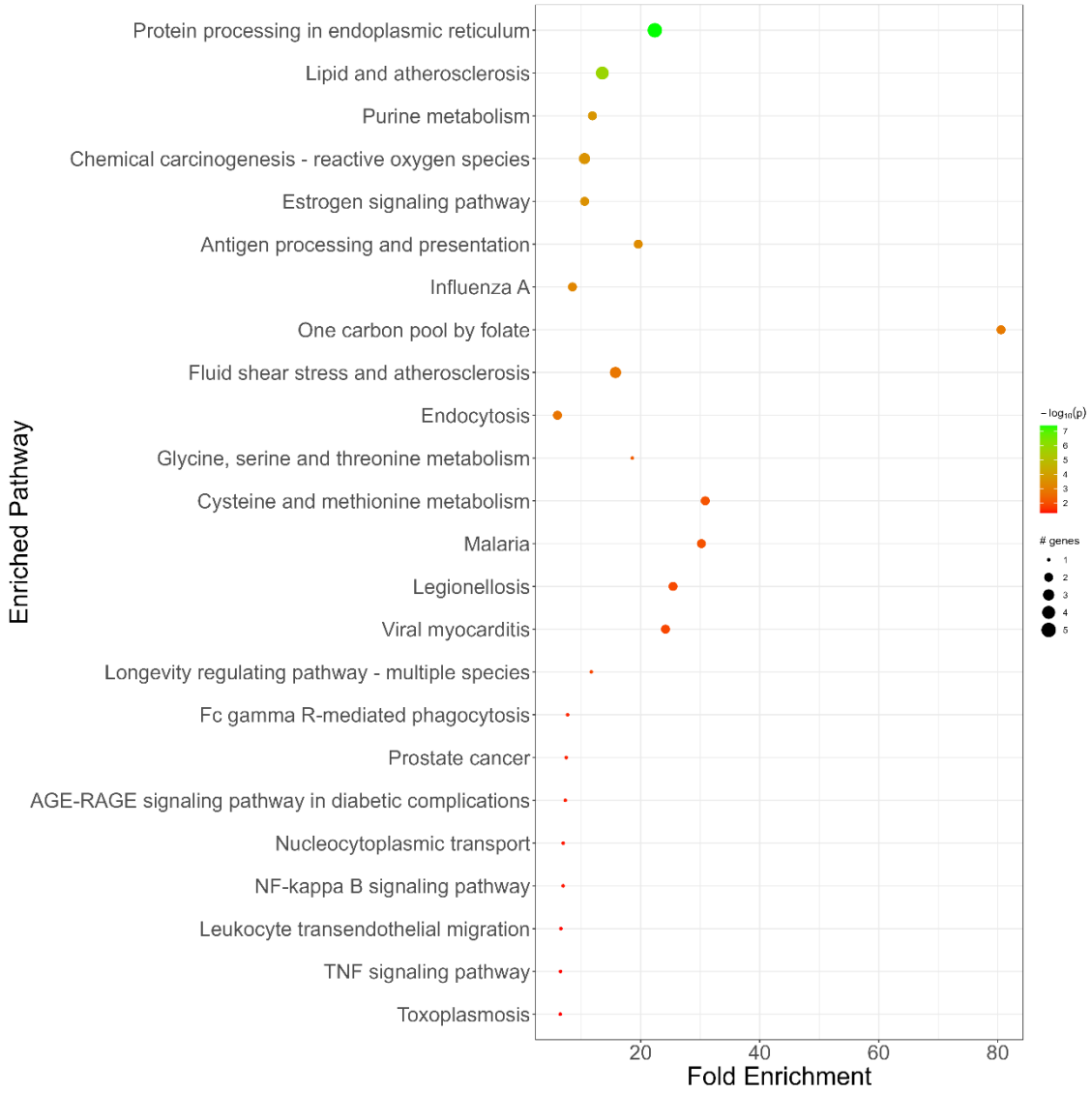


Figure 7.18: **Identification of all enriched pathways common to both CHO K1 and CHO DXB11 following adaptation to growth in serum free media. The colour of the bubble represents the p value associated with that pathway, while the size of the bubble represents the number of DE proteins identified in that pathway by pathfindR analysis. Some pathways are common to both categories hence why only fifteen are plotted.**

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated	Down Regulated
Top 10 KEGG paths ordered by significance					
hsa04141	Protein processing in endoplasmic reticulum	22.4	0.0002	HSPA8, HSP90AB1	CKAP4, RRBP1, DNAJC10
hsa00230	Purine metabolism	11.9	0.0002	GART, ATIC	None
hsa05164	Influenza A	8.53	0.0006	ICAM1, KPNA2	None
hsa05208	Chemical carcinogenesis - reactive oxygen species	10.6	0.0010	None	NDUFS1, EPHX1, NQO1
hsa05417	Lipid and atherosclerosis	13.5	0.0012	ICAM1, HSPD1, HSPA8, HSP90AB1	None
hsa05418	Fluid shear stress and atherosclerosis	15.8	0.0015	HSP90AB1, ICAM1	NQO1
hsa04144	Endocytosis	6.02	0.0017	HSPA8	EHD2
hsa04915	Estrogen signalling pathway	10.6	0.0026	HSP90AB1, HSPA8	None
hsa00260	Glycine, serine and threonine metabolism	18.6	0.0059	PSAT1	None
hsa00270	Cysteine and methionine metabolism	30.8	0.0085	GSS, PSAT1	None
Top 10 KEGG paths ordered by fold enrichment					
hsa04141	Protein processing in endoplasmic reticulum	22.4	0.0002	HSPA8, HSP90AB1	CKAP4, RRBP1, DNAJC10
hsa00230	Purine metabolism	11.9	0.0002	GART, ATIC	None
hsa05164	Influenza A	8.53	0.0006	ICAM1, KPNA2	None
hsa05208	Chemical carcinogenesis - reactive oxygen species	10.6	0.0010	None	NDUFS1, EPHX1, NQO1
hsa05417	Lipid and atherosclerosis	13.5	0.0012	ICAM1, HSPD1, HSPA8, HSP90AB1	None
hsa05418	Fluid shear stress and atherosclerosis	15.8	0.0015	HSP90AB1, ICAM1	NQO1
hsa04144	Endocytosis	6.02	0.0017	HSPA8	EHD2

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated	Down Regulated
hsa04915	Estrogen signalling pathway	10.6	0.0026	HSP90AB1, HSPA8	None
hsa00260	Glycine, serine and threonine metabolism	18.6	0.0059	PSAT1	None
hsa00270	Cysteine and methionine metabolism	30.8	0.0085	GSS, PSAT1	None

Table 7.11: **The ten KEGG pathways displaying the greatest degree of enrichment or significance following PSEA analysis of CHO K1 and CHO DXB11 cells adapted from growth in media containing serum to growth in serum free media. An additional criterion was applied in the generation of this table. PSEA analysis was performed on the DE proteins from K1 and DXB11 cells separately. This was because although the proteins were common to both cell lines the p value associated with them differed as the two cell lines were analysed in separate MS experiments. Therefore, only KEGG pathways identified as being enriched in both cell lines were reported in this table. For space reasons HGNC symbols rather than protein names have been used throughout this table. Protein names are listed in Table 7.14 at the end of this chapter.**

KEGG Path ID	Description	Fold Enrichment	p value	Up Regulated DE Proteins	Down Regulated DE Proteins
hsa04915	Estrogen signalling pathway	10.6	0.0026	HSP90AB1, HSPA8	None
Protein names are listed in Table 7.14 at the end of this chapter.					

Table 7.12: All adhesion related KEGG pathways found to be significantly enriched following adaptation of CHO K1 and CHO DXB11 cells from growth in media containing serum to growth in serum free media. For space reasons HGNC symbols rather than protein names have been used.

Only one KEGG pathway (Table 7.12) associated with adhesion were found to be statistically enriched during adaption of CHO K1 and DXB11 cells from growth in media containing serum to serum free media. This suggests that despite the data presented in Table7.7 and Table 7.8 adhesion is not the primary biological process being disrupted during the adaptation of CHO K1 and DXB11 cells to growth in serum free media.

Gene ontology analysis of the twenty differentially expressed proteins present in the enriched KEGG pathways revealed an association with one hundred and forty one different GO terms (Appendix 27). Table 7.13 lists the five most significantly enriched GO terms in each ontology associated with the enriched KEGG pathways identified by the PSEA analysis (Table 7.11).

ID	Description	p value	p adjusted	q value	# Proteins	Ontology
GO:0031667	response to nutrient levels	7.23E-06	0.004358	0.002654	6	BP
GO:0009991	response to extracellular stimulus	9.96E-06	0.004358	0.002654	6	BP
GO:0009410	response to xenobiotic stimulus	1.21E-05	0.004358	0.002654	4	BP
GO:0007568	Aging	1.73E-05	0.004358	0.002654	5	BP
GO:0042398	cellular modified amino acid biosynthetic process	1.99E-05	0.004358	0.002654	3	BP
GO:0016651	oxidoreductase activity, acting on NAD(P)H	5.02E-06	0.000613	0.000312	4	MF
GO:0015036	disulfide oxidoreductase activity	1.24E-05	0.000759	0.000386	3	MF
GO:0016667	oxidoreductase activity, acting on a sulfur group of donors	3.84E-05	0.00156	0.000794	3	MF
GO:0016668	oxidoreductase activity, acting on a sulfur group of donors, NAD(P) as acceptor	8.19E-05	0.002497	0.001271	2	MF
GO:0023026	MHC class II protein complex binding	0.000168	0.003419	0.00174	2	MF
GO:1904813	ficolin-1-rich granule lumen	0.000306	0.014113	0.008402	3	CC
GO:0044295	axonal growth cone	0.00041	0.014113	0.008402	2	CC
GO:0101031	chaperone complex	0.00041	0.014113	0.008402	2	CC
GO:0005765	lysosomal membrane	0.000638	0.014113	0.008402	4	CC
GO:0098852	lytic vacuole membrane	0.000638	0.014113	0.008402	4	CC

Table 7.13: **Significantly enriched GO terms following PSEA analysis of CHO K1 and CHO DXB11 cells adapted from growth in media containing serum to growth in serum free media. Adjusted p value and q Value are the p value following application of the Benjamini-Hochberg formula and false discovery rate (FDR) respectively to the calculated p value.**

As can been seen from Table 7.13, GO terms associated with molecular function dominate the list in terms of significance and suggest that the ability of cells to respond to oxidative stress via the generation of NADP or other oxidoreductase substrates is a key requirement under pining the adaption of CHO cells from suspension growth in media contain serum to serum free media. The list of most significantly enriched cellular components was dominated by ficolin granules. GO terms related to biological processes and cellular compartments such as cellular modified amino acid biosynthetic process, chaperone complex, lysosomal membrane and lytic vacuoles support the observation that protein processing in the endoplasmic reticulum was an enriched KEGG pathway and suggest that amino acid catabolism or the unfolded protein response maybe involved in the adaptation of CHO K1 and CHO DX11 cells to growth in serum free media.

7.4. Discussion

This study compared significantly ($p \leq 0.05$ and fold change ≥ 1.5) differentially expressed proteins detected in cells grown in media with and without serum. Two different cell lines CHO K1 and CHO DXB11 were used. Although the total number of differentially expressed proteins detected differed between CHO K1 and CHO DXB11 cell lines, one hundred and sixty-five differentially expressed proteins were determined to have a common profile in both CHO cell lines. A variety of bioinformatic tools were used to identify enriched KEGG pathways and gene ontology terms to identify an overall biological theme describing these common differentially expressed proteins. Considering that the difference between the various samples taken from each cell line was serum concentration, proteins/processes related to cell growth, the cell cycle, and cell metabolism were expected to be present. This expectation was based on the extensive body of literature describing multiple media formulations and their effect on viable cell concentration.

7.4.1. Reproducibility of the number of proteins identified in each replicate and the impact on the identification of enriched KEGG pathways

In excess of two thousand nine hundred proteins were identified in each replicate for all K1 serum concentrations. However, greater variability was observed across the DXB11 replicates. This variability was at its greatest in the 5% FBS, 1% FBS and 0.5% early serum concentrations where the number of proteins detected ranged from one thousand nine hundred and eighty two to four thousand two hundred and eight seven. Despite this variability only six proteins from the list of common enriched KEGG pathways (Table 7.5) were found to be not present in all sample replicates. These proteins and the five KEGG pathways they belong to are listed in Table 7.3 It's unlikely that the exclusion of *mitochondrial ribosomal protein S9* (ribosome)*, phosphoribosylformylglycinamidine synthase* (purine metabolism)*, scavenger receptor class B member 1* (cholesterol metabolism)*, VAMP associated protein A* (cholesterol metabolism) or *SIL1 nucleotide exchange factor* (protein processing in the endoplasmic reticulum) from the PSEA analysis will have any impact on the identification of these KEGG paths as enriched given the fact that multiple other DE proteins contributed to the identification of these KEGG pathways as being enriched in this dataset.

Given that only two proteins, farnesyl diphosphate synthase and 3-hydroxy-3-methylglutaryl-CoA synthase 2 were identified in this dataset as playing a role in the terpenoid backbone biosynthesis KEGG path the same justification can't be applied. However, as 3-hydroxy-3-methylglutaryl-CoA synthase 2 was detected in at least three replicates of each serum concentration for both the K1 and DXB11 cell line it was considered that this was a true differentially expressed protein and that the identification of the terpenoid backbone biosynthesis pathway as enriched was appropriate.

7.4.2. Glycolytic proteins are differentially expressed following adaption to serum free media.

The pathway by which one molecule of glucose is converted into two molecules of pyruvate is one of the processes by which the cell generates its energy and is known as glycolysis. The glycolytic enzymes *hexokinase*, *aldolase fructose bisphosphate A,*

phosphoglycerate kinase 1, enolase and *pyruvate kinase* were detected in our data. All these enzymes were shown to be upregulated in this dataset. These findings paint a picture where pyruvate potentially accumulates in the cell via the upregulation of glycolytic enzymes. This could theoretically result in an inability of cell to completely oxidase pyruvate resulting in the accumulation of acetaldehyde. Acetaldehyde can be oxidised to acetate by the enzyme aldehyde dehydrogenase 2 (Enomoto et al., 1991). Acetate, in turn, can be converted to acetyl CoA, an intermediary in the TCA cycle, the primary source of ATP production by cells. This process could be thought of as an overflow pathway for ensuring that excess pyruvate is directed to the TCA cycle rather than lactate production. *Aldehyde dehydrogenase 2 family member* was detected in our data but, in contrast to the other glycolytic enzymes present in our data, its expression was downregulated thus denying this overflow pathway to CHO cells adapted to growth in serum free media.

As well as converting phosphoenolpyruvate to pyruvate the upregulation of pyruvate kinase has a second negative effect in CHO cells. It has been shown by Neermann et al. (1996) that overexpression of pyruvate kinase results in the inhibition of pyruvate dehydrogenase, an enzyme which transports pyruvate into the mitochondria where it's converted into acetyl-CoA which then enters the TCA cycle.

While discussing energy generation by CHO cells attention is drawn to inositol 1,4,5-triphosphate receptor type 3 (ITPR3). Although ITPR3 is one of the three calcium trafficking channels found in the cell (reviewed by Mangla et al., 2020) and not a glycolytic enzyme it's well established that the ER and mitochondria are intracellular storage depots for Ca^{2+} ions. In response to external stimuli, the calcium receptors ITPR3 and ryanodine (RYR) become activated and mediate the movement of Ca^{2+} from the ER to the cytosol. This event results in regions of increased Ca^{2+} concentration adjacent to the mitochondria (Clapham, 2007), which stimulate the mitochondria to adsorb Ca^{2+}. The flow of Ca^{2+} from the cytosol into the mitochondria is a regulator of energy generation by the cell via the interaction of Ca^{2+} and TCA dehydrogenases (reviewed by Duchen, 2000 and McCormack et al., 1990). ITPR3 was downregulated in this dataset providing further evidence that it's via anaerobic fermentation that CHO cells derive their energy requirements.

7.4.3. Cholesterol metabolic/biosynthesis pathways are enriched in CHO K1 and DXB11 cells during the adaptation to growth in serum free media

The data presented in this study provides evidence for cholesterol metabolism playing a role in the adaptation of CHO K1 and DXB11 cells from growth in media containing serum to serum free media. *Neutral cholesterol ester hydrolase 1 (NCEH1), Scavenger receptor class B member 1 (SCARB1), Prolow-density lipoprotein receptor-related protein 1 (LRP1)* and *vesicle -associated membrane protein-associated protein A (VAPA)* are all DE proteins which are members of the cholesterol metabolism KEGG pathway and were found to be down regulated in this study. These proteins have diverse functions corresponding to the many functions cholesterol plays within the cell.

In the context of this study, SCARB1, LRP1 and VAPA are the most interesting. In addition to mediating the uptake into the cell of HDL, class B scavenger receptors (SRB1) have been shown (Fukasawa et al., 1996) to be present in CHO K1 cells where they recognised and mediated the endocytosis of cells undergoing apoptosis. Adaptation of CHO K1 and DXB11 cells to serum free media is characterised by periods of low culture viability as cell which fail to adapt die off. Perhaps the fact that this study found that SCARBI is downregulated is one of the mechanisms were by cells avoid apoptosis and as they adapt to changes in the growth media. In support of this, heat shock protein family A (Hsp70) member 8 (HSPA8), a member of the longevity regulating pathway, was shown to be upregulated in this study. Knockdown of HSPA8 expression with siRNAs had been shown to increase the apoptosis rate in endometrial cancer cell lines (Shan et al., 2016). Additionally, Lee et al. (2009), Ishaque et al. (2007) and Lasunskaia et al. (2003) used overexpression of HSP70 as a strategy to prolong culture viability of CHO and BHK cells as well as NS0 hybridomas by improving resistance to apoptosis. Prolow- density lipoprotein receptor-related protein 1 (LRP1) has also been implicated in the endocytosis and phagocytosis of apoptotic cells. Fouët et al. (2020) demonstrated using HEK293F cells that LRP1 specifically interacts with complement C1q which leads to the phagocytosis of apoptotic cells. LRP1 was found to be downregulated in this study. Vesicle -associated membrane protein-associated protein A has been shown to associate with homologs (ORPs) of oxysterol binding protein. The ORP3-VAPA complex has been shown to significantly increase β1-integrin activation in HEK293 cells relative

to controls (Weber-Boyvat et al., 2015). The same authors proposed that this association of ORP3 and VAPA could also enhance cellular survival signals.

Terpenoids are carbon-based molecules that are members of the terpene family. They are primarily isomeric hydrocarbons described by the molecular formula $(C_5H_8)_n$ (also known as terpenes). Over time the classification expanded to include the oxygenated, hydrogenated, and dehydrogenated derivatives of these hydrocarbon chains. Common terpenoids include isopentenyl pyrophosphate, a precursor of isopentenyl adenine (tRNA), retinol (vitamin A1), and farnesol. Farnesol is a precursor of squalene, the building block for all steroids and a key intermediary in cholesterol synthesis. All terpenoids are synthesised from 3 acetyl-CoA via the mevalonic acid pathway, reviewed by Andersen et al. (1990). *Isopentenyl diphosphate delta isomerase (IDI1), 3-hydroxy-3-methylglutaryl-CoA synthase 2 (HMGCS2) and farnesyl diphosphate synthase (FDPS)* are key enzymes in this pathway and were detected as being differentially expressed in this study. HMGCS2 catalyses the first step in the mevalonic acid pathway, the dehydration of acetyl CoA and acetoacetyl CoA to hydroxymethylglutarate CoA, IPP isomerase catalyses the conversion of isopentenyl pyrophosphate (IPP) to dimethylallyl pyrophosphate, and FDPS catalyses the conversion of geranyl pyrophosphate to farnesyl pyrophosphate, the progenitor of squalene

The literature reviewing terpenoids in mammalian cells is limited. In their 2018 review of the biochemistry of terpenoids, Ashour et al. (2018) discussed the role that terpenoids play in the post translational modification of proteins. They state, "in fact, all eukaryotic cells appear to contain proteins that have been post translationally modified by the attachment of C_{15} and C_{20} terpenoid side chains via a thioether linkage." Zhang and Casey (1996) demonstrated that the post translational modification of proteins in this manner increases the hydrophobicity of the protein and acts to target proteins to the cell membrane or control protein-protein interactions.

7.4.4. Cholesterol and the Warburg Effect

Cholesterol has many roles in the cell. It's an essential structural component of cell membranes, a precursor of steroid hormones, vitamin D and has a role in protein secretion (Ridsdale et al., 2006 and Kockx et al., 2012) and has been demonstrated to

be essential for growth in CHO cells (Xu et al., 2005). Cells can synthesise it internally or obtain it by internalising lipoproteins present in growth media. Historically the predominant source of cholesterol was via low density lipoproteins present in serum. Given that this study investigated what cellular pathways were perturbed following the adaptation to serum free media, it's not surprising that pathways that allowed the cell to synthesise cholesterol as a replacement for the lack of low-density lipoproteins in the media dominate the list.

However, there's another potential consequence for this upregulation of cholesterol biosynthesis. It's well documented that CHO cells suffer from the Warburg effect. This effect is characterised by the cell preferentially obtaining its energy requirements from anaerobic glycolysis rather than the aerobic TCA cycle and is characterised in cancer cells by increased cholesterogenesis (Siperstein and Fagan, 1964). This results in pyruvate being converted into lactate, which is then exported from the cell and consequently the media pH decreases. This has implications for culture osmolality as base must be added to maintain pH at the desired level. Elevated osmolality has been correlated with reduced cell culture viability (Zhu et al., 2008), increased levels of the undesirable glycoform Man 5 (Pacis et al., 2011) and decreased specific cell growth rate (Alhuthali et al., 2021) in CHO cells. Cell line and media development programmes have been devoted to minimising (Buchsteiner et al., 2017) or eliminating this undesirable phenomenon. Indeed, the Lewis group at the University of California San Diego reported the generation of CHO cells which didn’t exhibit the Warburg effect using a CRISPR/Cas9 gene engineering approach to knock out the enzymes responsible for lactate production and ancillary regulators, Hefzi et al. (2018) and Lewis et al. (2018). Cholesterol biosynthesis has been linked with the Warburg effect in cancer cells by multiple groups (reviewed by Coleman and Parlo, 2021). Therefore, the question is whether the development of serum free media and the consequent enrichment of cholesterol biosynthesis pathways, selected for cells that exhibited the Warburg effect phenotype?

This study provides some evidence for this hypothesis. The fact that enzymes involved in cholesterol biosynthesis are upregulated in addition to the observation that pyruvate kinase activity was upregulated, thus increasing the amount of pyruvate being produced by the cell via glycolysis. There's also evidence for aberrant pyruvate oxidation due to the presence of aldehyde dehydrogenase 2, an enzyme that breaks down the by-product

of incomplete pyruvate oxidation. Hence, all the pyruvate generated by the cell may not be entering the TCA cycle and could instead be converted to lactate. Gao et al. (2019) demonstrated using ALDH2 deficient mice that the intracellular accumulation of acetaldehyde had an inhibitory effect on glycolysis, thus reducing energy generation via glycolysis. This suggests that these cells derive their energy from anaerobic pathways.

The glucagon signalling pathway (reviewed by Jiang and Zhang, 2003) was found to be enriched in CHO K1 cells but not DXB11 cells. This pathway functions as an additional control mechanism for regulating glucose levels in the cell. Binding of glucagon to its receptor results in the activation of coupled G-proteins. One of these, G_q, results in the activation of phospholipase C and the production of inositol 1,4,5 triphosphate leading to the release of intracellular calcium. This causes decreased glycolysis, decreased glycogenesis but increased gluconeogenesis. The overall effect is increased amounts of glucose are available to the cell.

7.4.5. Impact of correcting for false discovery rates (FDR) on the identification of differentially expressed proteins common to CHO K1 and CHO DXB11 cells following adaptation to suspension growth in serum free media.

Interrogation of the DE protein list generated by Progenesis QI for Proteomics following application of a retrospective FDR criteria at the DE peptide selection step in Progenesis QI for Proteomics and an assessment of the impact omitting such a criteria had in the identification of DE proteins resulted in a reduction in the overall number of DE proteins in each cell line and in the number of DE proteins common to both CHO K1 and CHO DXB11 cell lines. While the overall picture presented by analysis of enriched KEGG pathways showed that metabolic pathways dominated in both scenarios there was a difference in the type of metabolic pathways which were found to be enriched. In the absence of an FDR filter during peptide selection the metabolism of carbohydrates and amino acids dominated. The application of an FDR filter resulted in KEGG pathways which were involved in amino acid metabolism dominating and the identification of a KEGG pathway involving the cellular response to reactive oxygen species. The DE proteins (*NADH ubiquinone oxidoreductase core subunit S1* and *NAD(P)H quinone dehydrogenase 1)* which were found to be part of this pathway support the idea that combating cell stress underpin the adaptation of cells to growth in serum free media.

7.4.5.1. CHO cells develop mechanisms to alleviate oxidative stress during the adaptation to growth in serum free media

Although the list of the most significant DE proteins by pvalue and fold change changed following the application of an FDR filter, the list of enriched KEGG pathways didn't change dramatically. Pathways such as protein processing in the endoplasmic reticulum, purine metabolism, the estrogen signalling pathway, one carbon pool by folate and cholesterol metabolism were common to both lists. A notable addition to the list of KEGG pathways following the application of the FDR filter was chemical carcinogenesis – reactive oxygen species and the down regulation of the enzymes *NADH ubiquinone oxidoreductase core subunit S1, epoxide hydrolase 1* and *NAD(P)H quinone dehydrogenase 1*. One of the benefits of including serum in media was to utilise its protective antioxidant properties. Halliwell (2003) mentions that serum contains some antioxidant enzymes, including low levels of catalase, superoxide dismutase, and glutathione peroxidase. Therefore, it makes sense that, as antioxidants, these enzymes bestowed a protective benefit on cells from exposure to ROS. It's therefore logical that following the adaptation of cells to serum free media enzymes such as NDUFS1 and NQO1 which are involved in the generation of ROS would be downregulated as cells look for methods to replace the protective effect of serum with respect to ROS.

This observation builds on the observation in Section 6.4.1.1 that one of the adaptations CHO K1 and DXB11 cells made during the adaptation from attached to suspension growth in the presence of serum was to upregulate the expression of NNT, an enzyme which regulates mitochondrial NADPH levels and oxidative stress by controlling the production of ROS within the cell.

7.5. Relevance of this study to CHO cell culture:

This chapter compared the proteome of CHO K1 and DXB11 cells as they adapted to growth in serum free media. It has provided insights into the biological pathways that potentially underscored the development of CHO cell lines capable of growth in serum free media. This knowledge can be helpful as the biopharma industry develops the serum-free processes of the future, for example, the growing trend for lab grown meat. In support of this suggestion Levitt et al. (2020) reported that the down regulation of

glycolytic enzymes observed in this study has also been observed in a study comparing myoblasts and myotubes.

7.6. DE proteins implicated in enriched KEGG pathways discussed in this chapter

HGNC Symbol	Protein Name	Uniprot ID	Listed In
ADK	adenosine kinase	P55263	Table 7.5
ALDH2	aldehyde dehydrogenase 2 family member	P05091	Table 7.5
ALDOA	aldolase, fructose-bisphosphate A	P04075	Table 7.5
ATIC	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	P31939	Table 7.5 and Table 7.11
CFL1	cofilin 1	P23528	Table 7.5 and Table 7.9
CKAP4	cytoskeleton associated protein 4	Q07065	Table 7.5 and Table 7.11
DNAJC10	DnaJ heat shock protein family (Hsp40) member C10	Q8IXB1	Table 7.5 and Table 7.11
DUT	deoxyuridine triphosphatase	P33316	Table 7.5
EDEM3	ER degradation enhancing alpha-mannosidase like protein 3	Q9BZQ6	Table 7.5
EHD2	EH domain containing 2	Q9NZN4	Table 7.11
ENO1	enolase 1	P06733	Table 7.5
EPHX1	epoxide hydrolase 1	P07099	Table 7.11
FDPS	farnesyl diphosphate synthase	P14324	Table 7.5
FKBP4	FKBP prolyl isomerase 4	Q02790	Table 7.9
GART	phosphoribosylglycinamide formyltransferase,	P22102	Table 7.5 and Table 7.11
GSN	Gelsolin	P06396	Table 7.5 and Table 7.9

HGNC Symbol	Protein Name	Uniprot ID	Listed In
GSS	glutathione synthetase	P48637	Table 7.11
HK1	hexokinase 1	P19367	Table 7.5
HMGCS2	3-hydroxy-3-methylglutaryl-CoA synthase 2	P54868	Table 7.5
HMOX1	heme oxygenase 1	P09601	Table 7.5
HPRT1	hypoxanthine phosphoribosyltransferase 1	P00492	Table 7.5
HSPA8	heat shock protein family A (Hsp70) member 8	P11142	Table 7.5, Table 7.9 and Table 7.11
HSPD1	heat shock protein family D (Hsp60) member 1	P10809	Table 7.11
HSP90AB1	heat shock protein 90 alpha family class B member 1	P08238	Table 7.5, Table 7.9 and Table 7.11
ICAM1	intercellular adhesion molecule 1	P05362	Table 7.11
ITPR3	inositol 1,4,5-trisphosphate receptor type 3	Q14573	Table 7.9
IQGAP1	IQ motif containing GTPase activating protein 1	P46940	Table 7.5 and Table 7.9
IQGAP2	IQ motif containing GTPase activating protein 2	Q13576	Table 7.5 and Table 7.9
KPNA2	karyopherin subunit alpha 2	P52292	Table 7.11
LRP1	LDL receptor related protein 1	Q07954	Table 7.5
NDUFS1	NADH:ubiquinone oxidoreductase core subunit S1	P28331	Table 7.11
MRPS9	mitochondrial ribosomal protein S9	P82933	Table 7.5
MTHFD1	methylenetetrahydrofolate dehydrogenase, cyclohydrolase and	P11586	Table 7.5
NQO1	NAD(P)H quinone dehydrogenase 1	P15559	Table 7.11

HGNC Symbol	Protein Name	Uniprot ID	Listed In
NCEH1	neutral cholesterol ester hydrolase 1	Q6PIU2	Table 7.5
NME2	NME/NM23 nucleoside diphosphate kinase 2	P22392	Table 7.5
PAICS	phosphoribosylaminoimidazole carboxylase and	P22234	Table 7.5
PCNA	proliferating cell nuclear antigen	P12004	Table 7.5
PDIA6	protein disulfide isomerase family A member 6	Q15084	Table 7.5
PFAS	phosphoribosylformylglycinamidine synthase	O15067	Table 7.5
PGK1	phosphoglycerate kinase 1	P00558	Table 7.5
PKM	pyruvate kinase M1/2	P14618	Table 7.5
PNP	purine nucleoside phosphorylase	P00491	Table 7.5
PRPS1	phosphoribosyl pyrophosphate synthetase 1	P60891	Table 7.5
PSAT1	phosphoserine aminotransferase 1	Q9Y617	Table 7.5
PSMA6	proteasome 20S subunit alpha 6	P60900	Table 7.5
PSMB3	proteasome 20S subunit beta 3	P49720	Table 7.5
PSMB5	proteasome 20S subunit beta 5	P28074	Table 7.5
PSMC1	proteasome 26S subunit, ATPase 1	P62191	Table 7.5
PSMC3	proteasome 26S subunit, ATPase 3	P17980	Table 7.5
PSMC6	proteasome 26S subunit, ATPase 6	P62333	Table 7.5
PSMD6	proteasome 26S subunit, non-ATPase 6	Q15008	Table 7.5

HGNC Symbol	Protein Name	Uniprot ID	Listed In
RAD23B	RAD23 homolog B, nucleotide excision repair protein	P54727	Table 7.5
RDX	Radixin	P35241	Table 7.5 and Table 7.9
RPA1	replication protein A1	P27694	Table 7.5
RPL4	ribosomal protein L4	P36578	Table 7.5
RPL5	ribosomal protein L5	P46777	Table 7.5
RPL18	ribosomal protein L18	Q07020	Table 7.5
RRBP1	ribosome binding protein 1	Q9P2E9	Table 7.5 and Table 7.11
RRM1	ribonucleotide reductase catalytic subunit M1	P23921	Table 7.5
SCARB1	scavenger receptor class B member 1	Q8WTV0	Table 7.5
SIL1	Q9H173	Q9H173	Table 7.5
VAPA	VAMP associated protein A	Q9P0L0	Table 7.5

Table 7.14: **Protein names associated with the HGNC symbols listed in Table 7.5, Table 7.9, and Table 7.11.**

CHAPTER 8

Concluding Remarks and Future Work

The work performed for this thesis ultimately aimed to develop a scale independent primary recovery process. The rationale behind this was that the development of a one size fits all scales primary recovery process would deliver a homogenous reproducible feed stream for downstream purification steps. This would help eliminate the challenges encountered, during downstream operations, while scaling process to the volumes necessary to support a commercial programme. Ultimately the development of a one size fits all scales primary recovery process could help reduce the cost of goods, a key metric for consideration when bringing biologics to market. Three different approaches were investigated based around flocculation and developing an approach which didn't require the use of synthetic flocculating agents such as pDADMAC.

The first approach undertaken investigated the use of fungal, yeast and human proteins as flocculating agents. A literature review identified suitable candidates but ultimately it was not possible to confirm the expression of these proteins in CHO K1 or CHO S cells. This chapter of the thesis was always recognised by the sponsor as having a high potential risk of failure but its novel approach (the use of Hydrophobins) made attempting it worthwhile. Advances in the scientific community's knowledge, since this thesis was started, of how to overcome the challenges associated with difficult to express proteins as discussed in chapter three open up further avenues for investigation with this approach. Approaches that have been exploited by other groups to improve expression levels include gene expression profiling combined with gene engineering (Lax et al., 2013), targeting transcriptionally active "hot spots" using site specific recombinases (Phan et al., 2017) or CRISPR/Cas9 (Kawabe et al., 2019) and using proteasome inhibitors (Knight et al., 2021). These approaches improved expression levels of the protein of interest by factors ranging from 3 fold (Knight et al., 2021) to 18 fold (Phan et al., 2017). Targeting transcriptionally active hotspots in conjunction with creating stable cell lines would allow for extended culture time compared to the transient transfection systems employed. It can't be ruled out that a critical concentration of hydrophobins is required to mediate flocculation and that transient transfections didn't give sufficient time for this concentration to be exceeded.

Knight et al. (2021) hypothesized that those cells with a greater capacity for protein turnover would have higher recombinant protein productivity. Application of a selective pressure via the inclusion of proteasome inhibitors in the culture media lead to

increased expression of a difficult to express fusion protein. While they don't elucidate the reason why this is so, one of the mechanisms they proposed to explain this was that inclusion of proteasome inhibitors selected for cells with an inherent folding capacity, reducing the throughput into the ERAD pathway and cellular stress due to reduced accumulation of misfolded proteins. There is a possibility that the apparent inability of CHO cells to express Hydrophobins is because they are classic difficult to express proteins. Hence the lack of expression could be because they are expressed as misfolded proteins and therefore targeted for degradation by the cell. Given the results reported by Knight et al. (2021) with their difficult to express protein inhibiting the proteasome seems to be an option worth exploring.

While the use of cell free expression systems would remove the need to perform a cell harvesting/removal step prior to commencing purification operations and as such was not an option, there may be value in using a CHO cell free expression system to investigate the question of whether CHO cells can express Hydrophobins since *E. coli* cell free expression systems have been successfully used in the production of Hydrophobins. However, obviously the use of cell free expression systems is in itself, a tool for overcoming the challenge posed when clarifying industrial processes with high viable cell concentrations at harvest.

Having investigated a cell engineering approach an investigation into the use of inorganic cations as flocculating agents was presented. Different cations at multiple concentrations were added to CHO K1 cells and the reduction in the number of cells/mL remaining in suspension was monitored. This part of the thesis was impacted by the Covid-19 pandemic. Ideally the data presented in this chapter would have formed a screening study and then a DOE approach combining the different cations investigated and multiple CHO cell lines would have been undertaken. This would have increased the statistical power of the study and potentially allowed the identification of novel combinations of inorganic cations which promote cellular flocculation and don't have a detrimental effect on the quality of the recombinant biologic. This approach would have also help further refine the relationship between the amount of cation required for flocculation and cell concentration.

Given cellular clumping was observed with all the cations investigated in this study it may also be interesting to perform a proteomic DE analysis comparing cells resulting

from cation concentrations which promoted clumped cells compared to concentrations which didn't. This approach has the potential to identify endogenous CHO proteins whose expression profile changes upon clumping. These would then be obvious targets for a cell engineering approach.

Further work could also be done with regards to investigating the impact using inorganic cations as flocculants had on key product quality attributes. In particular an investigation into the level of oxidative damage due to Fenton and Fenton like reactions resulting from the use of iron as a flocculating agent would be of benefit. Although cysteine, lysine and histidine can be targeted by oxidising agents, it tends to be methionine and to a lesser extent tryptophan residues which are subjected to oxidation when considering biologics and in particular antibodies. Both Haberger et al. (2015) and Mo et al. (2016) have published MS based approaches for the assessment of oxidative damage to antibodies. The advantage of using a native MS approach as described by Haberger et al. (2016) is that it could also act as an orthogonal method for providing information on the level of aggregation and antibody dimer formation.

Finally, a proteomic approach was investigated. This part of the thesis compared CHO K1 cells which had been preadapted to suspension growth in serum free media as attached and suspension cultures in the presence and absence of serum. Additionally, attached CHO K1 and CHO DXB11 cells which had not been adapted to suspension growth in serum free media were sequentially adapted to suspension growth and then growth in serum free media. This approach highlighted proteins and KEGG pathways common to both cells lines whose expression profile changed in line with changes in culture format and media composition. Given that membrane proteins and phosphorylated proteins such as cadherins and focal adhesion kinase are known to play a central role in cellular adhesion it's surprising that these haven't been identified in the DE protein lists identified in these chapters. A potential explanation for this may be found in the experimental approach utilised. The approach taken in this thesis focussed on proteins extracted from whole cell lysates. It may be worthwhile investigating if including membrane enrichment or phosphoprotein enrichment techniques such as that described by Coleman et al. (2018) and Kaushik et al. (2018) as part of the sample preparation increases the number of DE proteins identified utilising a DE approach. Additionally, it would be useful to experimentally verify by either western blotting, qPCR

or RNA sequencing if the proteins identified in chapters 6 and 7 as being part of adhesion related KEGG pathways were differentially expressed in clumping and non-clumping cell lines.

It may also be helpful to perform this study in the reverse order, i.e., adapt attached cells to serum free media followed by adaptation to suspension. By then comparing the two sets of common DE proteins resulting from the adaptation to serum free media to each other, proteins whose expression changes in response to the removal of serum from the media independent of culture format could be identified. Likewise, proteomic changes in the response to changes in culture format which are independent of serum could be identified by comparing the two sets of common DE proteins resulting from the adaptation to suspension growth.

The strategy employed in the adaptation of CHO cells to growth in serum free media was a stepwise one over a prolonged period. Samples were taken and analysed by mass spectrometry at each serum concentration and hence by modifying the well established approach whereby spent media analysis has been used to inform media optimisation studies (Liu et al., 2015) it would be possible to construct a profile of the expression profile of individual proteins over time and link changes in a specific protein or group of proteins expression to specific serum concentrations. Such information would be useful as industry moves to designing culture media or adaptation studies for novel cell culture processes of the future.

Something that wasn't explored as part of this thesis but is a key component of a flocculating cell line is how would flocculation be induced. Perhaps combining a proteomics and transcriptomic approaches would identify novel genetic tools such micro RNAs which could be exploited to achieve this aim. In their 2011 paper, Valastyan and Weinberg reviewed the role micro RNAs play as regulators of cell adhesion. The paper discusses a range of micro RNAs which target proteins involved cell-cell adhesion, cell-ECM adhesion proteins and cytoskeletal regulatory proteins. An alternative approach to developing an inducible expression system is the use of temperature as the inducer. Cooling of bioreactor contents prior to primary recovery is a common strategy in industrial facilities. Identification of a protein whose expression is induced at below physiological temperatures would be the first step in designing such as system. Cold shock proteins are the obvious candidates for this role. A cold shock protein CSDE1, *Cold*

shock domain containing E1, was identified as being downregulated during the adaptation of CHO K1 and CHO DXB11 cells to suspension (Chapter 6). Further work could verify if this protein was expressed during bioreactor cooling and if so could its upregulation promote cellular flocculation.

In summary, the projects undertaken for this thesis have taken the first steps towards identifying approaches which could be further exploited during the development of an inducible flocculating cell line for use in the biologics industry.

BIBLIOGRAPHY

Aceto, N., et al. (2014) 'Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis', *Cell,* Vol. 158, pp. 1110-1122.

Alhuthali, S., Kotidis, P. and Kontoravdi, C. (2021) 'Osmolality Effects on CHO Cell Growth, Cell Volume, Antibody Productivity and Glycosylation', *International Journal of Molecular Sciences*, 22(7), pp. 3290. doi:10.3390/ijms22073290

Allison, D.S. (1999) 'Hamster EF-1α transcriptional regulatory DNA'. *Available at: https://patents.google.com/patent/US5888809A/en* (Accessed: 2 April 2022).

Alves, C. S. and Dobrowsky, T. M. (2017) 'Heterologous Protein Production in CHO Cells' Springer Science+Business Media LLC (Methods in Molecular Biology). Available at: doi:10.1007/978-1-4939-6972-2.

Andersen, R.J. (1990) 'Terpenoids from Selected Marine Invertebrates', in Towers, G.H.N. and Stafford, H.A. (eds) *Biochemistry of the Mevalonic Acid Pathway to Terpenoids*. Boston, MA: Springer US, pp. 265–282. doi:10.1007/978-1-4684-8789-3_8.

Arakawa, M. (2010) 'Inhibition of Monocyte Adhesion to Endothelial Cells and Attenuation of Atherosclerotic Lesion by a Glucagon-like Peptide-1 Receptor Agonist, Exendin-4', *Diabetes*, 59(4), pp. 1030–1037. doi:10.2337/db09-1694.

Ashburner, M. et al. (2000) 'Gene Ontology: tool for the unification of biology', *Nature Genetics*, 25(1), pp. 25–29. doi.org/10.1038/75556.

Ashour, M., Wink, M. and Gershenzon, J. (2018) 'Biochemistry of Terpenoids: Monoterpenes, Sesquiterpenes and Diterpenes', in *Annual Plant Reviews online. American Cancer Society*, pp. 258–303. doi:10.1002/9781119312994.apr0427.

Aubin, J., Davy, A. and Soriano, P. (2004) 'In vivo convergence of BMP and MAPK signaling pathways: impact of differential Smad1 phosphorylation on development and homeostasis', *Genes & Development*, 18(12), pp. 1482–1494. doi.org/10.1101/gad.1202604.

Aunins, J.G. and Wang, D.I.C. (1989) 'Induced flocculation of animal cells in suspension culture', *Biotechnology and Bioengineering*, 34(5), pp. 629–638. doi:10.1002/bit.260340507.

Bachmann, M. (2019) 'Cell Adhesion by Integrins', *Physiological Reviews*, 99(4), pp. 1655–1699. doi:10.1152/physrev.00036.2018.

Baumgartner-Parzer, S.M. et al. (1995) ‘Modulation by high glucose of adhesion molecule expression in cultured endothelial cells’, *Diabetologia*, 38: pp 1367-1370. doi:10.1007/BF00401771

Baycin-Hizal, D. et al. (2012) ‘Proteomic Analysis of Chinese Hamster Ovary Cells’, *Journal of Proteome Research*, 11(11), pp. 5265–5276. doi:10.1021/pr300476w.

Bebbington, C.R. et al. (1992) ‘High-Level Expression of a Recombinant Antibody from Myeloma Cells Using a Glutamine Synthetase Gene as an Amplifiable Selectable Marker’, *Bio/Technology*, 10(2), pp. 169–175. doi:10.1038/nbt0292-169.

Benjamini, Y. and Hochber, Y. (1995) ‘Controlling the false discovery rate: A practical and powerful approach to multiple testing’ J. Royal Statistical Society: Series B, (Methodological) 57(1) pp. 289-300. Doi:10.1111/j.2517-6161.1995.tb02031.x

Besson, A. and Guiraud, P. (2013) ‘High-pH-induced flocculation–flotation of the hypersaline microalga Dunaliella salina’, *Bioresource Technology*, 147, pp. 464–470. doi.org/10.1016/j.biortech.2013.08.053.

Bertrand, V. et al. (2019) ‘Transcriptome and proteome analysis of steady-state in a perfusion CHO cell culture process’, *Biotechnology and Bioengineering*, 116(8), pp. 1959–1972. doi:10.1002/bit.26996.

Beynon, R.J. and Pratt, J.M. (2005) ‘Metabolic Labelling of Proteins for Proteomics’, *Molecular & Cellular Proteomics*, 4(7), pp. 857–872. doi:10.1074/mcp.R400010-MCP200.

Bianchi, A.A. and McGrew, J.T. (2003) ‘High-level expression of full-length antibodies using trans-complementing expression vectors’, *Biotechnology and Bioengineering*, 84(4), pp. 439–444. doi:10.1002/bit.10790.

Bidard, F. et al. (1994) ‘Cloning and analysis of a FLO5 flocculation gene from S. cerevisiae’, *Current Genetics*, 25(3), pp. 196–201. doi:10.1007/BF00357162.

Boraston, R. et al. (1992) ‘Elimination of cell aggregation in suspension cultures of Chinese hamster ovary (CHO) cells’, in Spier, R.E., Griffiths, J.B., and MacDonald, C. (eds) *Animal Cell Technology,* pp. 424–426. doi:10.1016/B978-0-7506-0421-5.50099-4.

Bornhorst, J. et al. (2012) ‘Effects of manganese and arsenic species on the level of energy related nucleotides in human cells’, *Metallomics*, 4(3), pp. 297–306. doi:10.1039/c2mt00164k.

Bort, J.A.H. et al. (2012) 'Dynamic mRNA and miRNA profiling of CHO-K1 suspension cell cultures', *Biotechnology Journal*, 7(4), pp. 500–515. doi.org/10.1002/biot.201100143.

Brahimi-Horn, M.C. and Pouysségur, J. (2009) 'HIF at a glance', *Journal of Cell Science,* 122(8), pp. 1055–1057. doi:10.1242/jcs.035022.

Branco, L.M. (2014) 'Rapid Generation of Stable NS0 Production Cell Lines in Chemically Defined Medium', *BioProcess International, 1 May*. Available at: https://bioprocessintl.com/2014/rapid-generation-of-stable-ns0-production-cell-lines-in-chemically-defined-medium-351600/ (Accessed: 2 April 2022).

Bryan, L. et al. (2021) 'Mapping the molecular basis for growth related phenotypes in industrial producer CHO cell lines using differential proteomic analysis', *BMC Biotechnology*, 21(1), p. 43. doi.org/10.1186/s12896-021-00704-8.

Buchsteiner, M. et al. (2018) 'Improving culture performance and antibody production in CHO cell culture processes by reducing the Warburg effect', *Biotechnology and Bioengineering*, 115(9), pp. 2315–2327. doi:10.1002/bit.26724.

Burgstaller, D. et al. (2018) 'Continuous cell flocculation for recombinant antibody harvesting', *Journal of Chemical Technology & Biotechnology*, 93(7), pp. 1881–1890. doi:10.1002/jctb.5500.

Capella Roca, B. et al. (2019) 'An arginase-based system for selection of transfected CHO cells without the use of toxic chemicals', *Journal of Biological Chemistry*, 294(49), pp. 18756–18768. doi:10.1074/jbc.RA119.011162.

Cartwright, J.F. et al. (2020) 'A platform for context-specific genetic engineering of recombinant protein production by CHO cells', *Journal of Biotechnology*, 312, pp. 11–22. doi:10.1016/j.jbiotec.2020.02.012.

Cary, L.A., Chang, J.F. and Guan, J.-L. (1996) 'Stimulation of cell migration by overexpression of focal adhesion kinase and its association with Src and Fyn', *Journal of Cell Science*, Vol. 109, No. 7, pp. 1787-1794.

Cha, S. et al. (2010) 'In Situ Proteomic Analysis of Human Breast Cancer Epithelial Cells Using Laser Capture Microdissection: Annotation by Protein Set Enrichment Analysis and Gene Ontology', *Molecular & Cellular Proteomics*, 9(11), pp. 2529–2544. doi:10.1074/mcp.M110.000398.

Chapman D. L. (1913) 'A contribution to the theory of electrocapillary', Philos Mag, 25:475 481.

Chen, G. et al. (2017) 'Application of calcium phosphate flocculation in high-density cell culture fluid with high product titer of monoclonal antibody', *Bioprocess and Biosystems Engineering*, 40(5), pp. 703–714. doi:10.1007/s00449-017-1735-9.
Cheng, S.Y.S. et al. (2014) 'Grb2 Promotes Integrin-Induced Focal Adhesion Kinase (FAK) Autophosphorylation and Directs the Phosphorylation of Protein Tyrosine Phosphatase α by the Src-FAK Kinase Complex', *Molecular and Cellular Biology*, 34(3), pp. 348–361. doi:10.1128/MCB.00825-13.
Chevallier, V. et al. (2020) 'Characterization of glutathione proteome in CHO cells and its relationship with productivity and cholesterol synthesis', *Biotechnology and Bioengineering*, 117(11), pp. 3448–3458. doi:10.1002/bit.27495.
Clapham, D.E. (2007) 'Calcium Signaling', *Cell*, 131(6), pp. 1047–1058. doi:10.1016/j.cell.2007.11.028.
Coleman, O. et al. (2017) 'Filter-Aided Sample Preparation (FASP) for Improved Proteome Analysis of Recombinant Chinese Hamster Ovary Cells', in Meleady, P. (ed.) *Heterologous Protein Production in CHO Cells: Methods and Protocols.* New York, NY: Springer (Methods in Molecular Biology), pp. 187–194. doi:10.1007/978-1-4939-6972-2_12.
Coleman, O. et al. (2018) 'A Comparative Quantitative LC-MS/MS Profiling Analysis of Human Pancreatic Adenocarcinoma, Adjacent-Normal Tissue, and Patient-Derived Tumour Xenografts', *Proteomes*, 6(4), p. 45. doi:10.3390/proteomes6040045.
Coleman, P.S. and Parlo, R.A. (2021) 'Warburg's Ghost—Cancer's Self-Sustaining Phenotype: The Aberrant Carbon Flux in Cholesterol-Enriched Tumor Mitochondria via Deregulated Cholesterogenesis', *Frontiers in Cell and Developmental Biology*, 9, p. 278. doi:10.3389/fcell.2021.626316.
Conway, J.R., Lex, A. and Gehlenborg, N. (2017) 'UpSetR: an R package for the visualization of intersecting sets and their properties', *Bioinformatics*, 33(18), pp. 2938–2940. doi.org/10.1093/bioinformatics/btx364.
Cooper, G., M. (2000) *The Cell: A Molecular Approach. 2nd edition*. Sinauer Associates 2000.
Csermely, P. et al. (1998) 'The 90-kDa Molecular Chaperone Family: Structure, Function, and Clinical Applications. A Comprehensive Review', *Pharmacology & Therapeutics*, 79(2), pp. 129–168. doi:10.1016/S0163-7258(98)00013-8.

Cui, L. et al. (2020) ‘Excretory overexpression of hydrophobins as multifunctional biosurfactants in E. coli’, *International Journal of Biological Macromolecules*, 165, pp. 1296–1302. doi:10.1016/j.ijbiomac.2020.09.206.

Damluji, R. and Riley, P. (1979) ‘On the role of calcium in adhesion of cells to solid substrates.’, *Exp. Cell Biol.*, 47(3), pp. 226–37. doi.org/10.1159/000162940.

de Vocht, M.L. et al. (2000) ‘Structural and Functional Role of the Disulfide Bridges in the Hydrophobin SC3’, *Journal of Biological Chemistry*, 275(37), pp. 28428–28432. doi:10.1074/jbc.M000691200.

de Vries, O.M.H. et al. (1993) ‘Insoluble hydrophobin complexes in the walls of Schizophyllum commune and other filamentous fungi’, *Archives of Microbiology*, 159(4), pp. 330–335. doi:10.1007/BF00290915.

Demir, I. et al. (2020) ‘Towards a better understanding of microalgae natural flocculation mechanisms to enhance flotation harvesting efficiency’, *Water Science and Technology*, 82(6), pp. 1009–1024. doi.org/10.2166/wst.2020.177.

Dennis G. et al. (2003) ‘DAVID: Database for Annotation, Visualization, and Integrated Discovery.’ *Genome Biol.*, 4(5), R60. doi.org/10.1186/gb-2003-4-9-r60

Domingo-Almenara, X. et al. (2019) ‘Autonomous METLIN-Guided In-source Fragment Annotation for Untargeted Metabolomics’, *Analytical Chemistry*, 91(5), pp. 3246–3253. doi:10.1021/acs.analchem.8b03126.

Duchen, M.R. (2000) ‘Mitochondria and calcium: from cell signalling to cell death’, *The Journal of Physiology*, 529(Pt 1), pp. 57–68. doi:10.1111/j.1469-7793.2000.00057.x.

Dumont, J. et al. (2016) ‘Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives’, *Critical Reviews in Biotechnology*, 36(6), pp. 1110–1122. doi.org/10.3109/07388551.2015.1084266.

Durinck, S. et al. (2005) ‘BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis’, *Bioinformatics (Oxford, England)*, 21(16), pp. 3439–3440. doi.org/10.1093/bioinformatics/bti525.

Endo, Y. and Alonso, Manuel. (2001) ‘Physical Meaning of Specific Cake Resistance and Effects of Cake Properties in Compressible Cake Filtration’, *Filtration & Separation*, 38(7), pp. 42–46. doi:10.1016/S0015-1882(01)80447-X.

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.' *J Am Soc Mass Spectrom,* 5(11), pp. 976-89. doi: 10.1016/1044-0305(94)80016-2.

Enomoto, N. et al. (1991) 'Acetaldehyde Metabolism in Different Aldehyde Dehydrogenase-2 Genotypes', *Alcohol Clin Exp Res*, 15(1), pp. 141–144.

Fanelli, M.A. et al. (2008) 'P-Cadherin and β-catenin are useful prognostic markers in breast cancer patients; β-catenin interacts with heat shock protein Hsp27', *Cell Stress and Chaperones*, 13(2), pp. 207–220. doi:10.1007/s12192-007-0007-z.

Feary, M. et al. (2021) 'CHOK1SV GS-KO SSI expression system: A combination of the Fer1L4 locus and glutamine synthetase selection', *Biotechnology Progress*, 37(4), p. e3137. doi:10.1002/btpr.3137.

Feichtinger, J. et al. (2016) 'Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time', *Biotechnology and Bioengineering*, 113(10), pp. 2241–2253. doi:10.1002/bit.25990.

Forrest, M.E. et al. (2020) 'Codon and amino acid content are associated with mRNA stability in mammalian cells', *PLOS ONE*, 15(2), p. e0228730. doi:10.1371/journal.pone.0228730.

Fouët, G. et al. (2020) 'Complement C1q Interacts With LRP1 Clusters II and IV Through a Site Close but Different From the Binding Site of Its C1r and C1s-Associated Proteases', *Frontiers in Immunology*, 11. Article 583754. doi.org/10.3389/fimmu.2020.583754

Francisco, L.F.V. et al. (2021) 'Acute Toxic and Genotoxic Effects of Aluminum and Manganese Using in Vitro Models', *Toxics*, 9(7), p. 153. doi:10.3390/toxics9070153.

Freeberg, M.A.T. et al. (2018) 'Serpine1 Knockdown Enhances MMP Activity after Flexor Tendon Injury in Mice: Implications for Adhesions Therapy', *Scientific Reports*, 8(1), p. 5810. doi:10.1038/s41598-018-24144-1.

Frenkel, H.S. and Van Waveren, G.M. (1938) 'Cultivation to Foot, and Mouth Disease Virus on Foetal Bovine and Ovine Skin.', Cultivation to Foot, and Mouth Disease Virus on Foetal Bovine and Ovine Skin. [Preprint]. *Available at: https://www.cabdirect.org/cabdirect/abstract/19412201453* (Accessed: 24 January 2021).

Froud, S.J. (1999) 'The development benefits and disadvantages of serum free media', *Dev. Biol. Stand.*, 99, pp. 157–66.

Fukasawa, M. et al. (1996) 'SRB1, a Class B Scavenger Receptor, Recognizes both Negatively Charged Liposomes and Apoptotic Cells', *Experimental Cell Research*, 222(1), pp. 246–250. doi.org/10.1006/excr.1996.0030.

Ganendren, R. et al. (2006) 'Phospholipase B activity enhances adhesion of Cryptococcus neoformans to a human lung epithelial cell line', *Microbes and Infection*, 8(4), pp. 1006–1015. doi.org/10.1016/j.micinf.2005.10.018.

Gao, Y. et al. (2019) 'Alcohol inhibits T-cell glucose metabolism and hepatitis in ALDH2-deficient mice and humans: roles of acetaldehyde and glucocorticoids', *Gut*, 68(7), pp. 1311–1322. doi:10.1136/gutjnl-2018-316221.

Gasser, F., Mulsant, P. and Gillois, M. (1985) 'Long-term multiplication of the Chinese hamster ovary (CHO) cell line in a serum-free medium', *In Vitro Cellular amp; Developmental Biology*, 21(10), pp. 588–592. doi:10.1007/BF02620890.

Giansanti, P. et al. (2016) 'Six alternative proteases for mass spectrometry–based proteomics beyond trypsin', *Nature Protocols*, 11(5), pp. 993–1006. doi.org/10.1038/nprot.2016.057.

Glover, Z.K. et al. (2015) 'Metal ion interactions with mAbs: Part 1', *mAbs*, 7(5), pp. 901–911. doi:10.1080/19420862.2015.1062193.

Godoy-Parejo, C. et al. (2019) 'Insulin Stimulates PI3K/AKT and Cell Adhesion to Promote the Survival of Individualized Human Embryonic Stem Cells', *STEM CELLS*, 37(8), pp. 1030–1041. doi.org/10.1002/stem.3026.

Goossens, K. and Willaert, R. (2010) 'Flocculation protein structure and cell–cell adhesion mechanism in Saccharomyces cerevisiae', *Biotechnology Letters*, 32(11), pp. 1571–1585. doi.org/10.1007/s10529-010-0352-3.

Goldfarb, A.R., Saidel, L.J. and Mosovich, E. (1951) 'THE ULTRAVIOLET ABSORPTION SPECTRA OF PROTEINS'*, Journal of Biological Chemistry*, 193(1), pp. 397–404. doi:10.1016/S0021-9258(19)52465-6.

Grigorieva, J. et al. (2020) 'Application of protein set enrichment analysis to correlation of protein functional sets with mass spectral features and multivariate proteomic tests', *Clinical Mass Spectrometry*, 15, pp. 44–53. doi:10.1016/j.clinms.2019.09.001.

Gouy M. (1910) 'Sur la constitution de la charge électrique a la surface d'un électrolyte', *J. Phys*, 9 :457-468.

Guo, B. et al. (2000) 'A Saccharomyces gene family involved in invasive growth, cell–cell adhesion, and mating', *Proceedings of the National Academy of Sciences*, 97(22), pp. 12158–12163. doi:10.1073/pnas.220420397.

Gutierrez, J.M. et al. (2020) 'Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion', *Nature Communications*, 11(1), p. 68. doi.org/10.1038/s41467-019-13867-y.

Gutteridge, J.M.C., Maidt, L. and Poyer, L. (1990) 'Superoxide dismutase and Fenton chemistry. Reaction of ferric-EDTA complex and ferric-bipyridyl complex with hydrogen peroxide without the apparent formation of iron(II)', *Biochemical Journal*, 269(1), pp. 169–174. doi:10.1042/bj2690169.

Hacker, D.L., De Jesus, M. and Wurm, F.M. (2009) '25 years of recombinant proteins from reactor-grown cells — Where do we go from here?', *Biotechnology Advances*, 27(6), pp. 1023–1027. doi: 10.1016/j.biotechadv.2009.05.008

Hägg, K., Cimbritz, M. and Persson, K.M. (2018) 'Combining Chemical Flocculation and Disc Filtration with Managed Aquifer Recharge', *Water*, 10(12), p. 1854. doi.org/10.3390/w10121854.

Halbleib, J.M. and Nelson, W.J. (2006) 'Cadherins in development: cell adhesion, sorting, and tissue morphogenesis', *Genes & Development*, 20(23), pp. 3199–3214. doi:10.1101/gad.1486806.

Halliwell, B. (2003) 'Oxidative stress in cell culture: an under-appreciated problem?' *FEBS Letters*, 540(27106), pp. 3-6

Hamilton, G. W. and Ham, R.G. (1977) 'Clonal growth of Chinese hamster cell lines in protein-free media', *In Vitro*, 13(9), pp. 537–547. doi:10.1007/BF02627849.

Hausmann, R. et al. (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*, 265, pp. 65–69. doi:10.1016/j.jbiotec.2017.11.008.

Haworth, A.S. and Brackenbury, W.J. (2019) 'Emerging roles for multifunctional ion channel auxiliary subunits in cancer', *Cell Calcium*, 80, pp. 125–140. doi.org/10.1016/j.ceca.2019.04.005.

Hawk M. A., and Schafer T. Z. (2018) 'Mechanisms of redox metabolism and cancer cell survival during extracellular matrix detachment', *J. Biol. Chem*., 293, pp. 7531-7537.

Hawk, M.A. et al. (2018) 'RIPK1-mediated induction of mitophagy compromises the viability of extracellular-matrix-detached cells', *Nature Cell Biology*, 20(3), pp. 272–284. doi.org/10.1038/s41556-018-0034-2.

Heffner, K.M. et al. (2017) 'Lessons from the Hamster: Cricetulus griseus Tissue and CHO Cell Line Proteome Comparison', *Journal of Proteome Research*, 16(10), pp. 3672–3687. doi:10.1021/acs.jproteome.7b00382.

Heffner, K. et al. (2020) 'Expanded Chinese hamster organ and cell line proteomics profiling reveals tissue-specific functionalities', *Scientific Reports*, 10(1), p. 15841. doi:10.1038/s41598-020-72959-8.

Hefzi, H. et al. (2018) 'Elimination of the Warburg effect in Chinese hamster ovary (CHO) cells improves cell phenotype as a protein production platform', *in Cell Culture Engineering XVI*. A. Robinson, PhD, Tulane University R. Venkat, PhD, MedImmune E. Schaefer, ScD, J&J Janssen Eds, (ECI Symposium Series). Available at: https://dc.engconfintl.org/ccexvi/45.

Heinrich, C.A. et al. (1999) 'Negative regulation of N-cadherin-mediated cell-cell adhesion by the estrogen receptor signaling pathway in rat pituitary GH3 cells', *Endocrine*, 10(1), pp. 67–76. doi:10.1385/ENDO:10:1:67.

Hellemans, J. and Vandesompele, J. (2011) *qPCR data analysis – unlocking the secret to successful results*. Caister Academic Press.

Henry, M. et al. (2017) 'Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift', *Journal of Proteome Research*, 16(7), pp. 2339–2358. doi:10.1021/acs.jproteome.6b00868.

Hill, R.P., Marie-Egyptienne, D.T. and Hedley, D.W. (2009) 'Cancer Stem Cells, Hypoxia and Metastasis', *Seminars in Radiation Oncology*, 19(2), pp. 106–111. doi:10.1016/j.semradonc.2008.12.002.

Hu, J. et al. (2019) 'Epidermal growth factor-containing fibulin-like extracellular matrix protein 1 (EFEMP1) suppressed the growth of hepatocellular carcinoma cells by promoting Semaphorin 3B(SEMA3B)', *Cancer Medicine*, 8(6), pp. 3152–3166. doi:10.1002/cam4.2144.

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

Hu, Y.-Y. et al. (2012) 'Notch signalling pathway and cancer metastasis', *Advances in Experimental Medicine and Biology*, 727, pp. 186–198. doi:10.1007/978-1-4614-0899-4_14.

Hutterer, K.M. et al. (2013) 'Monoclonal antibody disulfide reduction during manufacturing', *mAbs*, 5(4), pp. 608–613. doi:10.4161/mabs.24725.

Ishaque, A. et al. (2007) 'Over-expression of Hsp70 in BHK-21 cells engineered to produce recombinant factor VIII promotes resistance to apoptosis and enhances secretion', *Biotechnology and Bioengineering*, 97(1), pp. 144–155. doi:10.1002/bit.21201.

Isserlin, R. et al. (2010) 'Pathway analysis of dilated cardiomyopathy using global proteomic profiling and enrichment maps', *PROTEOMICS*, 10(6), pp. 1316–1327. doi:10.1002/pmic.200900412.

Jaber, J. et al. (2011) 'Stimulus responsive polymers for the purification of biomolecules'. Available at: *https://patents.google.com/patent/US20110313066A1/en (*Accessed: 10 April 2022).

Janse van Rensburg, H.J. and Yang, X. (2016) 'The roles of the Hippo pathway in cancer metastasis', *Cellular Signalling*, 28(11), pp. 1761–1772. doi.org/10.1016/j.cellsig.2016.08.004.

Janiszewska, M., Primi, M.C. and Izard, T. (2020) 'Cell adhesion in cancer: Beyond the migration of single cells', *Journal of Biological Chemistry*, 295(8), pp. 2495–2505. doi:10.1074/jbc.REV119.007759.

Jayme, D., Watanabe, T. and Shimada, T. (1997) 'Basal medium development for serum-free culture: a historical perspective', *Cytotechnology* 23(1-3): pp. 95–101. doi: 10.1023/A:1007967602484.

Jiang, G. and Zhang, B.B. (2003) 'Glucagon and regulation of glucose metabolism', *American Journal of Physiology-Endocrinology and Metabolism*, 284(4), pp. E671–E678. doi.org/10.1152/ajpendo.00492.2002.

Jiang, L. et al. (2016) 'Reductive carboxylation supports redox homeostasis during anchorage-independent growth', *Nature*, 532(7598), pp. 255–258. doi.org/10.1038/nature17393.

Jones, D. et al. (2003) 'High-Level Expression of Recombinant IgG in the Human Cell Line PER.C6', *Biotechnology Progress*, 19(1), pp. 163–168. doi:10.1021/bp025574h.

Jung, K. (2016). '*Statistical Analysis in Proteomics - False Discovery Rate Estimation in Proteomics'*, Method's in Molecular Biology (Chapter 7), 1362, 119–128. doi:10.1007/978-1-4939-3106-4_7.

Kanehisa, M. and Goto, S. (2000) 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Research*, 28(1), pp. 27–30.

Kanehisa, M. et al. (2014) 'Data, information, knowledge and principle: back to metabolism in KEGG', *Nucleic Acids Research*, 42(D1), pp. D199–D205. doi:10.1093/nar/gkt1076.

Kang, Y. (Kenneth) et al. (2013) 'Development of a novel and efficient cell culture flocculation process using a stimulus responsive polymer to streamline antibody purification processes', *Biotechnology and Bioengineering*, 110(11), pp. 2928–2937. doi:10.1002/bit.24969.

Kaur, G. et al. (2017) 'Affected energy metabolism under manganese stress governs cellular toxicity', *Scientific Reports*, 7(1), p. 11645. doi:10.1038/s41598-017-12004-3.

Kaushik, P. et al. (2018) 'The Expression Pattern of the Phosphoproteome Is Significantly Changed During the Growth Phases of Recombinant CHO Cell Culture', *Biotechnology Journal*, 13(10), p. 1700221. doi:10.1002/biot.201700221.

Kawabe, Y. et al. (2018) 'Targeted knock-in of an scFv-Fc antibody gene into the hprt locus of Chinese hamster ovary cells using CRISPR/Cas9 and CRIS-PITCh systems', *Journal of Bioscience and Bioengineering*, 125(5), pp. 599–605. doi:10.1016/j.jbiosc.2017.12.003.

Kelly, S (2016) *Proteomic characterisation of clonal populations from a human lung carcinoma cell line.* PhD thesis, Dublin City University.

Keszey, Z. and Sütö, Z. (2020) 'Improvement of affinity chromatography of immunoglobulins by using pre-capture flocculation'. *Available at: https://patents.google.com/patent/WO2020200980A1/en* (Accessed: 28 May 2022).

Kim, Y.-B. et al. (2005) ‘Cell Adhesion Status-dependent Histone Acetylation Is Regulated through Intracellular Contractility-related Signaling Activities’, *Journal of Biological Chemistry*, 280(31), pp. 28357–28364. doi:10.1074/jbc.M412608200.

Kisko, K. et al. (2008) ‘Interactions of Hydrophobin Proteins in Solution Studied by Small-Angle X-Ray Scattering’, *Biophysical Journal*, 94(1), pp. 198–206. doi:10.1529/biophysj.107.112359.

Klingler, F. et al. (2021) ‘Unveiling the CHO surfaceome: Identification of cell surface proteins reveals cell aggregation-relevant mechanisms’, *Biotechnology and Bioengineering*, 118(8), pp. 3015–3028. doi:10.1002/bit.27811.

Knight, T.J. et al. (2021) ‘Selection of CHO host and recombinant cell pools by inhibition of the proteasome results in enhanced product yields and cell specific productivity’, *Journal of Biotechnology*, 337, pp. 35–45. doi:10.1016/j.jbiotec.2021.06.019.

Kockx, M. et al. (2012) ‘Cholesterol accumulation inhibits ER to Golgi transport and protein secretion: studies of apolipoprotein E and VSVGt’, *Biochemical Journal*, 447(1), pp. 51–60. doi:10.1042/BJ20111891.

Krężel, A. et al. (2010) ‘Sequence-Specific Ni(II)-Dependent Peptide Bond Hydrolysis for Protein Engineering. Combinatorial Library Determination of Optimal Sequences’, *Journal of the American Chemical Society*, 132(10), pp. 3355–3366. https://doi.org/10.1021/ja907567r.

Kuczewski, M. et al. (2011) ‘A single-use purification process for the production of a monoclonal antibody produced in a PER.C6 human cell line’, *Biotechnology Journal*, 6(1), pp. 56–65. doi:10.1002/biot.201000292.

Kumari, S.S. and Varadaraj, K. (2009) ‘Intact AQP0 performs cell-to-cell adhesion’, *Biochemical and Biophysical Research Communications*, 390(3), pp. 1034–1039. doi:10.1016/j.bbrc.2009.10.103.

Kumar, S. et al. (2022) ‘A proteomics approach to decipher a sticky CHO situation’, *Biotechnology and Bioengineering*, 119(8), pp. 2064–2075. doi.org/10.1002/bit.28108.

Labuschagne, C. F. et al. (2019), ‘Cell Clustering Promotes a Metabolic Switch that Supports Metastatic Colonization’, *Cell Metabolism*, 30(4), pp.720-734. https://doi.org/10.1016/j.cmet.2019.07.014.

Lahtinen, T. et al. (2008) 'Hydrophobin (HFBI): A potential fusion partner for one-step purification of recombinant proteins from insect cells', *Protein Expression and Purification*, 59(1), pp. 18–24. doi:10.1016/j.pep.2007.12.014.

Lalonde, M.-E. and Durocher, Y. (2017) 'Therapeutic glycoprotein production in mammalian cells', *Journal of Biotechnology*, 251, pp. 128–140. doi:10.1016/j.jbiotec.2017.04.028.

Lasunskaia, E.B. et al. (2003) 'Transfection of NS0 myeloma fusion partner cells with HSP70 gene results in higher hybridoma yield by improving cellular resistance to apoptosis', *Biotechnology and Bioengineering*, 81(4), pp. 496–504. doi:10.1002/bit.10493.

Laux, H. et al. (2013) 'Generation of genetically engineered CHO cell lines to support the production of a difficult to express therapeutic protein', *BMC Proceedings*, 7(6), p. P1. doi:10.1186/1753-6561-7-S6-P1.

Le Boeuf, F., Houle, F. and Huot, J. (2004) 'Regulation of Vascular Endothelial Growth Factor Receptor 2-mediated Phosphorylation of Focal Adhesion Kinase by Heat Shock Protein 90 and Src Kinase Activities*', *Journal of Biological Chemistry*, 279(37), pp. 39175–39185. doi.org/10.1074/jbc.M405493200.

Lee, J.W. et al. (2019) 'Hypoxia signaling in human diseases and therapeutic targets', *Experimental & Molecular Medicine*, 51(6), pp. 1–13. doi:10.1038/s12276-019-0235-1.

Lee, N. et al. (2016) 'Targeted Gene Deletion Using DNA-Free RNA-Guided Cas9 Nuclease Accelerates Adaptation of CHO Cells to Suspension Culture', *ACS Synthetic Biology*, 5(11), pp. 1211–1219. doi:10.1021/acssynbio.5b00249.

Lee, Y.Y. et al. (2009) 'Overexpression of heat shock proteins (HSPs) in CHO cells for extended culture viability and improved recombinant protein production', *Journal of Biotechnology*, 143(1), pp. 34–43. doi:10.1016/j.jbiotec.2009.05.013.

LeMerdy, S. (2014) 'Evolving Clarification Strategies to Meet New Challenges', *BioProcess International*, 16 October. Available at: https://bioprocessintl.com/downstream-processing/filtration/evolving-clarification-strategies-meet-new-challenges/ (Accessed: 13 July 2021).

Lemieux, P. et al. (1997) 'The small heat shock protein hsp27 increases invasiveness but decreases motility of breast cancer cells', *Invasion & Metastasis*, 17(3), pp. 113–123.

Leng, C.H. et al. (2015) ‘Composite flocculants based on magnesium salt–polydiallyldimethylammonium chloride: characterization and flocculation behaviour’, *RSC Advances*, 5(66), pp. 53462–53470. doi:10.1039/C5RA03897A.
Levitt, D.E. et al. (2020) ‘Ethanol-Impaired Myogenic Differentiation is Associated with Decreased Myoblast Glycolytic Function’, *Alcoholism: Clinical and Experimental Research*, 44(11), pp. 2166–2176. doi:10.1111/acer.14453.
Lewis, C.W., Johnston, J.R. and Martin, P.A. (1976) ‘The Genetics of Yeast Flocculation’, *Journal of the Institute of Brewing*, 82(3), pp. 158–160. doi:10.1002/j.2050-0416.1976.tb03741.x.
Lewis, N.E. et al. (2013) ‘Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome’, *Nature Biotechnology*, 31(8), pp. 759–765. doi.org/10.1038/nbt.2624.
Lewis, N. et al. (2018) ‘Elimination of the “Essential” Warburg effect in CHO cells through a multiplex genome engineering strategy’, *Cell Culture Engineering XVI*
Lex, A. et al. (2014) ‘UpSet: Visualization of Intersecting Sets’, *IEEE Transactions on Visualization and Computer Graphics*, 20(12), pp. 1983–1992. doi:10.1109/TVCG.2014.2346248.
Li Tay, S., Yao Tan, H. and Perera, C. (2006) ‘The Coagulating Effects of Cations and Anions on Soy Protein’, *International Journal of Food Properties*, 9(2), pp. 317–323. doi:10.1080/10942910600596340.
Li, F. et al. (2010) ‘Cell culture processes for monoclonal antibody production’, *mAbs*, 2(5), pp. 466–479. doi:10.4161/mabs.2.5.12720.
Li, S. et al. (2018) ‘Proteogenomic annotation of the Chinese hamster reveals extensive novel translation events and endogenous retroviral elements’ *Bioinformatics*. doi:10.1101/468181.
Li S., et al. (2018) ‘Nicotinamide nucleotide transhydrogenase-mediated redox homeostasis promotes tumor growth and metastasis in gastric cancer’ *Redox Biol.* (18), pp. 246-255. doi: 10.1016/j.redox.2018.07.017
Li, W. et al. (2021) ‘Serum-Free Medium for Recombinant Protein Expression in Chinese Hamster Ovary Cells’, *Frontiers in Bioengineering and Biotechnology*, 9, p. 172. doi:10.3389/fbioe.2021.646363.

Livak, K.J. and Schmittgen, T.D. (2001) 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method', *Methods*, 25(4), pp. 402–408. doi.org/10.1006/meth.2001.1262.

Liu, W. et al. (2011) 'Tumor-derived vascular endothelial growth factor (VEGF)-a facilitates tumor metastasis through the VEGF-VEGFR1 signaling pathway', *International Journal of Oncology*, 39(5), pp. 1213–1220. doi:10.3892/ijo.2011.1138.

Liu, X. et al. (2011) 'Cell surface heat shock protein 90 modulates prostate cancer cell adhesion and invasion through the integrin-β1/ focal adhesion kinase/c-Src signaling pathway', *oncology reports*, p. 9.

Lu, W. et al. (2017) 'Metabolite Measurement: Pitfalls to Avoid and Practices to Follow', *Annual Review of Biochemistry*, 86(1), pp. 277–304. doi:10.1146/annurev-biochem-061516-044952.

Lualdi, M. and Fasano, M. (2019) 'Statistical analysis of proteomics data: A review on feature selection' *Journal of Proteomics*, 198, pp.18-26. doi.org/10.1016/j.jprot.2018.12.004.

Lucas, B.K. et al. (1996) 'High-Level Production of Recombinant Proteins in CHO Cells Using a Dicistronic DHFR Intron Expression Vector', *Nucleic Acids Research*, 24(9), pp. 1774–1779. doi:10.1093/nar/24.9.1774.

Lvovich Vadim, S.A. (1995) 'Complexation of Nicotinamide Adenine Dinucleotide with Ferric and Ferrous Ions', *Archives of Biochemistry and Biophysics*, 320(1), pp. 1-13. doi:10.1006/abbi.1995.1336.

Malm, M. et al. (2020) 'Evolution from adherent to suspension: systems biology of HEK293 cell line development', *Scientific Reports*, 10(1), p. 18996. doi:10.1038/s41598-020-76137-8.

Mangeat, P., Roy, C. and Martin, M. (1999) 'ERM proteins in cell adhesion and membrane dynamics', *Trends in Cell Biology*, 9(5), pp. 187–192. doi.org/10.1016/S0962-8924(99)01544-5.

Mangla, A., Guerra, M.T. and Nathanson, M.H. (2020) 'Type 3 inositol 1,4,5-trisphosphate receptor: A calcium channel for all seasons', *Cell calcium*, 85, p. 102132. doi:10.1016/j.ceca.2019.102132.

Mathews, M.T. and Berk, B.C. (2008) 'PARP-1 Inhibition Prevents Oxidative and Nitrosative Stress–Induced Endothelial Cell Death via Transactivation of the VEGF

Receptor 2', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 28(4), pp. 711–717. doi.org/10.1161/ATVBAHA.107.156406.

Mauro, V.P. and Chappell, S.A. (2014) 'A critical analysis of codon optimization in human therapeutics', *Trends in molecular medicine*, 20(11), pp. 604–613. doi.org/10.1016/j.molmed.2014.09.003.

McCormack, J.G., Halestrap, A.P. and Denton, R.M. (1990) 'Role of calcium ions in regulation of mammalian intramitochondrial metabolism', *Physiological Reviews*, 70(2), pp. 391–425. doi:10.1152/physrev.1990.70.2.391.

McNerney, T. et al. (2015) 'PDADMAC flocculation of Chinese hamster ovary cells: Enabling a centrifuge-less harvest process for monoclonal antibodies', *mAbs*, 7(2), pp. 413–427. doi:10.1080/19420862.2015.1007824.

Meleady, P. et al. (2011) 'Sustained productivity in recombinant Chinese Hamster Ovary (CHO) cell lines: proteome analysis of the molecular basis for a process-related phenotype', *BMC Biotechnology*, 11(1), pp. 78. doi:10.1186/1472-6750-11-78.

Meleady, P. et al. (2012a) 'Impact of miR-7 over-expression on the proteome of Chinese hamster ovary cells', *Journal of Biotechnology*, 160(3), pp. 251–262. doi.org/10.1016/j.jbiotec.2012.03.002.

Meleady, P. et al. (2012b) 'Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics', *Biotechnology and Bioengineering*, 109(6), pp. 1386–1394. doi.org/10.1002/bit.24476

Mihaylova M.M. and, Shaw R.J. (2011) 'The AMPK signalling pathway coordinates cell growth, autophagy and metabolism', *Nat Cell Biol. 2;* 13(9) pp.1016-23. doi: 10.1038

Miki, B.L.A. et al. (1982) 'Possible Mechanism for Flocculation Interactions Governed by Gene FLO1 in Saccharomyces cerevisiae', *Journal of Bacteriology*, Vol.150 (2), pp 878-889. doi: 10.1023/A:1007967602484

Milani AT, Khadem-Ansari MH, Rasmi Y. (2018) 'Effects of thyroid-stimulating hormone on adhesion molecules and pro-inflammatory cytokines secretion in human umbilical vein endothelial cells. - Abstract - Europe PMC', *Research in Pharmaceutical Sciences*, 13(6), pp. 546–56. doi:10.4103/1735-5362.245966.

Milani, A., Khadem-Ansari, M. and Rasmi, Y. (2019) 'Effects of thyroxine on adhesion molecules and proinflammatory cytokines secretion on human umbilical vein

endothelial cells', *Research in Pharmaceutical Sciences*, 14(3), p. 237. doi:10.4103/1735-5362.258490.

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

Moore, G. E., & Ulrich, K. (1965) 'Suspension cultures of mammalian cells.', *Journal of Surgical Research*, 5(6), pp. 270–82. doi:doi:10.1016/s0022-4804(65)80017-8.

Moreira, A.S. et al. (2020) 'Establishing Suspension Cell Cultures for Improved Manufacturing of Oncolytic Adenovirus', *Biotechnology Journal*, 15(4), p. 1900411. doi:https://doi.org/10.1002/biot.201900411.

Morris, V.K., Kwan, A.H. and Sunde, M. (2013) 'Analysis of the Structure and Conformational States of DewA Gives Insight into the Assembly of the Fungal Hydrophobins', *Journal of Molecular Biology*, 425(2), pp. 244–256. doi:10.1016/j.jmb.2012.10.021.

Murata, M., Eto, Y. and Shibai, H. (1988) 'Large-scale production of Erythroid Differentiation Factor (EDF) by gene-engineered Chinese hamster ovary (CHO) cells in suspension culture', *Journal of Fermentation Technology*, 66(5), pp. 501–507. doi:10.1016/0385-6380(88)90082-9.

Navarro-Tito, N., Robledo, T. and Salazar, E.P. (2008) 'Arachidonic acid promotes FAK activation and migration in MDA-MB-231 breast cancer cells', *Experimental Cell Research*, 314(18), pp. 3340–3355. doi.org/10.1016/j.yexcr.2008.08.018.

Neermann, J. and Wagner, R. (1996) 'Comparative analysis of glucose and glutamine metabolism in transformed mammalian cell lines, insect and primary liver cells', *Journal of Cellular Physiology*, 166(1), pp. 152–169. doi:10.1002/(SICI)1097-4652(199601)166:1<152::AID-JCP18>3.0.CO;2-H.

Neilson, K.A. et al. (2011) 'Less label, more free: Approaches in label-free quantitative mass spectrometry', *PROTEOMICS*, 11(4), pp. 535–553. doi:https://doi.org/10.1002/pmic.201000553.

Nissom, P.M. et al. (2006) 'Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells', *Molecular Biotechnology*, 34(2), pp. 125–140. doi:10.1385/MB:34:2:125.

Niu, B. et al. (2012) 'Heterologous expression and characterization of the hydrophobin HFBI in Pichia pastoris and evaluation of its contribution to the food industry', *Amino Acids*, 43(2), pp. 763–771. doi:10.1007/s00726-011-1126-5.
Ochel, H.-J. et al. (1999) 'The Benzoquinone Ansamycin Geldanamycin Stimulates Proteolytic Degradation of Focal Adhesion Kinase', *Molecular Genetics and Metabolism*, 66(1), pp. 24–30. doi.org/10.1006/mgme.1998.2774.
Oexle, H., Gnaiger, E. and Weiss, G. (1999) 'Iron-dependent changes in cellular energy metabolism: influence on citric acid cycle and oxidative phosphorylation', *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1413(3), pp. 99–107. doi:10.1016/S0005-2728(99)00088-2.
Oshima, T. et al. (2001) 'Regulation and distribution of MAdCAM-1 in endothelial cells in vitro', *American Journal of Physiology-Cell Physiology*, 281(4), pp. C1096–C1105. doi:10.1152/ajpcell.2001.281.4.C1096.
Ouellette, D. et al. (2009) 'Elevated cleavage of human immunoglobulin gamma molecules containing a lambda light chain mediated by iron and histidine', *Analytical Biochemistry*, 389(2), pp. 107–117. doi:10.1016/j.ab.2009.03.027.
Pacis, E. et al. (2011) 'Effects of cell culture conditions on antibody N-linked glycosylation—what affects high mannose 5 glycoform', *Biotechnology and Bioengineering*, 108(10), pp. 2348–2358. doi:10.1002/bit.23200.
Park, S.-Y. et al. (2021) 'Untargeted proteomics reveals upregulation of stress response pathways during CHO-based monoclonal antibody manufacturing process leading to disulfide bond reduction', *mAbs*, 13(1), p. 1963094. doi:10.1080/19420862.2021.1963094.
Patel, V.J. et al. (2009) 'A comparison of labeling and label-free mass spectrometry-based proteomics approaches', *Journal of Proteome Research*, 8(7), pp. 3752–3759. doi:10.1021/pr900080y.
Peng, Y. et al. (2002). 'Improved dataset characterisation for meta learning'. In International Conference on Discovery Science, pp. 141–152. Springer.
Perkins D.N. et al. (1999) 'Probability-based protein identification by searching sequence databases using mass spectrometry data.' *Electrophoresis*, 20(18), pp. 3551-67. doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

Peroutka, R.J. et al. (2008) 'Enhanced protein expression in mammalian cells using engineered SUMO fusions: Secreted phospholipase A2', Protein Science: *A Publication of the Protein Society*, 17(9), pp. 1586–1595. doi:10.1110/ps.035576.108.

Petan T., Jarc E. and Jusović M. (2018), 'Lipid Droplets in Cancer: Guardians of Fat in a Stressful World', *Molecules*, 23(8), pp 1941. doi: 10.3390/molecules23081941.

Petruzzelli, L., Takami, M. and Humes, D. (1999) 'Structure and Function of Cell Adhesion Molecules', *Am. J. Med.*, 106, pp. 467–476.

Pham, T.V. et al. (2012) 'Label-free mass spectrometry- based proteomics for biomarker discovery and validation', *Expert Rev. Mol. Diagn.*, 12(4), pp. 343–359

Phan, Q.V. et al. (2017) 'Site-specific chromosomal gene insertion: Flp recombinase versus Cas9 nuclease', *Scientific Reports*, 7, p. 17771. doi:10.1038/s41598-017-17651-0.

Pierschbacher, M.D. and Ruoslahti, E. (1984) 'Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule.', *Nature*, 309(5963), pp. 30–33. doi:10.1038/309030a0.

Pristovšek, N. et al. (2019) 'Systematic Evaluation of Site-Specific Recombinant Gene Expression for Programmable Mammalian Cell Engineering', *ACS Synthetic Biology*, 8(4), pp. 758–774. doi:10.1021/acssynbio.8b00453.

Puck, Cieciura and Robinson, T.T., S.J. and A.R. (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), pp. 945–956. doi:10.1084/jem.108.6.945.

Ramirez OT, Mutharasan R. (1992) 'Effect of serum on the plasma membrane fluidity of hybridomas: an insight into its shear protective mechanism.' *Biotechnol Prog.*, 8(1), pp 40-50. doi:10.1021/bp00013a007.

Rankin, E.B., Nam, J.-M. and Giaccia, A.J. (2016) 'Hypoxia: Signaling the Metastatic Cascade', *Trends in Cancer*, 2(6), pp. 295–304. doi: 10.1016/j.trecan.2016.05.006.

Rao K.N.S. et al. (2020) 'Nicotinamide nucleotide transhydrogenase (NNT) regulates mitochondrial ROS and endothelial dysfunction in response to angiotensin II', *Redox Biol.* 36(101650). doi: 10.1016/j.redox.2020.101650.

Reuten, R. et al. (2016) 'Maltose-Binding Protein (MBP), a Secretion-Enhancing Tag for Mammalian Protein Expression Systems', *PLOS ONE*, 11(3), p. e0152386. doi: 10.1371/journal.pone.0152386.

Reynolds, T.B. and Fink, G.R. (2001) 'Bakers' Yeast, a Model for Fungal Biofilm Formation', *Science*, 291(5505), pp. 878–881. doi:10.1126/science.291.5505.878.

Ridsdale, A. et al. (2006) 'Cholesterol Is Required for Efficient Endoplasmic Reticulum-to-Golgi Transport of Secretory Membrane Proteins', *Molecular Biology of the Cell*, 17(4), pp. 1593–1605. doi:10.1091/mbc. E05-02-0100.

Riske, F. et al. (2007) 'The use of chitosan as a flocculant in mammalian cell culture dramatically improves clarification throughput without adversely impacting monoclonal antibody recovery', *Journal of Biotechnology*, 128(4), pp. 813–823. doi.org/10.1016/j.jbiotec.2006.12.023.

Roche, J. and Törnroth-Horsefield, S. (2017) 'Aquaporin Protein-Protein Interactions', *International Journal of Molecular Sciences*, 18(11), p. 2255. doi.org/10.3390/ijms18112255.

Romero, J. et al. (2015) 'Method of isolating biomacromolecules using low pH and divalent cations'. Available at: *https://patents.google.com/patent/US9109015B2/en*.

Rousseau, S., Houle, F. and Huot, J. (2000) 'Integrating the VEGF Signals Leading to Actin-Based Motility in Vascular Endothelial Cells', *Trends in Cardiovascular Medicine*, 10(8), pp. 321–327. doi.org/10.1016/S1050-1738(01)00072-X.

Ross, D., Saxena, M. and Altmann, M. (2012) 'eIF4E Is an Important Determinant of Adhesion and Pseudohyphal Growth of the Yeast S. cerevisiae', *PLOS ONE*, 7(11). doi:10.1371/journal.pone.0050773.

Sandri, C. et al. (2012) 'The R-Ras/RIN2/Rab5 complex controls endothelial cell adhesion and morphogenesis via active integrin endocytosis and Rac signaling', *Cell Research*, 22(10), pp. 1479–1501. doi:10.1038/cr.2012.110.

Sawada, M. et al. (2002) 'Reduced sialidase expression in highly metastatic variants of mouse colon adenocarcinoma 26 and retardation of their metastatic ability by sialidase overexpression', *International Journal of Cancer*, 97(2), pp. 180–185. doi:10.1002/ijc.1598.

Schafer, Z.T., et al. (2009) 'Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment' *Nature*, 461, pp. 109-113.

Schrag, D., et al. (2014) 'Size Exclusion-High-Performance Liquid Chromatography (SEC-HPLC).' In: Ossipow, V., Fischer, N. (eds) Monoclonal Antibodies. Methods in Molecular Biology, vol 1131. Humana Press, Totowa, NJ. doi.org/10.1007/978-1-62703-992-5_31

Schratt, G. et al. (2002) 'Serum response factor is crucial for actin cytoskeletal organization and focal adhesion assembly in embryonic stem cells', *The Journal of Cell Biology*, 156(4), pp. 737–750. doi.org/10.1083/jcb.200106008.

Schröder, M. (2006) 'The unfolded protein response', *Molecular Biotechnology*, 34(2), pp. 279–290. doi:10.1385/MB:34:2:279.

Schuren, F.H.J. and Wessels, J.G.H. (1990) 'Two genes specifically expressed in fruiting dikaryons of Schizophyllum commune: homologies with a gene not regulated by mating-type genes', *Gene*, 90(2), pp. 199–205. doi:https://doi.org/10.1016/0378-1119(90)90180-Y.

Semerjian L and Ayoub G (2003) 'Magnesium chloride flocculant in the Wastewater treatment industry'. Available at: *https://www.researchgate.net/publication/223320692_High-pH-Magnesium_Coagulation-Flocculation_in_Wastewater_Treatment?enrichId=rgreq-743c19867e06d63508db6dd8be0e1413-XXX&enrichSource=Y292ZXJQYWdlOzIyMzMyMDY5MjtBUzo1NDg1NzQ5Njc0MTg4ODFAMTUwNzgwMTg3Mjk0MQ%3D%3D&el=1_x_3&_esc=publicationCoverPdf.*

Shan, N. et al. (2016) 'Identification of HSPA8 as a candidate biomarker for endometrial carcinoma by using iTRAQ-based proteomic analysis', *OncoTargets and Therapy*, 9, pp. 2169–2179. doi:10.2147/OTT.S97983.

Shen, C.F. et al. (2019) 'Development of suspension adapted Vero cell culture process technology for production of viral vaccines', *Vaccine,* 37(47), pp. 6996–7002. doi:10.1016/j.vaccine.2019.07.003.

Shiels, A. et al. (2001) 'Optical dysfunction of the crystalline lens in aquaporin-0-deficient mice', *Physiol. Genomics*, 7, pp. 179–186. doi: 10.1152/physiolgenomics.00078.2001.

Shimono, Y. et al. (2012) Immunoglobulin Superfamily Receptors and Adherens Junctions. Adherens Junctions: from *Molecular Mechanisms to Tissue Development and Disease*.

Shridhar, S. et al. (2017) 'Transcriptomic changes in CHO cells after adaptation to suspension growth in protein-free medium analysed by a species-specific microarray', *Journal of Biotechnology*, 257, pp. 13–21. doi:10.1016/j.jbiotec.2017.03.012.

Sicari, D. et al. (2020) ‘A guide to assessing endoplasmic reticulum homeostasis and stress in mammalian systems’, *The FEBS Journal*, 287(1), pp. 27–42. doi:10.1111/febs.15107.

Siddiquee, R. et al. (2020) ‘Cell-free expression of natively folded hydrophobins’, *Protein Expression and Purification*, 170, p. 105591. doi:10.1016/j.pep.2020.105591.

Sieg, D.J., Hauck, C.R. and Schlaepfer, D.D. (1999) ‘FAK promotes cell migration’, *Journal of Cell Science* Vol. 112 (16), pp 2677–2691.

Simpson, C. D., Anyiwe, K., and Schimmer A. D. (2008) **‘**Anoikis resistance and tumor metastasis’, *Cancer Lett.,* 272, pp. 177-185

Sinacore, M.S. et al. (1996) ‘CHO DUKX cell lineages preadapted to growth in serum-free suspension culture enable rapid development of cell culture processes for the manufacture of recombinant proteins’, *Biotechnology and Bioengineering*, 52(4), pp. 518–528. doi:https://doi.org/10.1002/(SICI)1097-0290(19961120)52:4<518::AID-BIT7>3.0.CO;2-S.

Sinacore, M.S., Drapeau, D. and Adamson, S.R. (2000) ‘Adaptation of mammalian cells to growth in serum-free media’, *MOLECULAR BIOTECHNOLOGY*, 15, pp. 9.

Singh, N. and Peck, M. (2014) ‘pDADMAC Flocculant Reagent for Use with Clarisolve® Depth Filters’, BioProcess International, 15 August. Available at: *https://bioprocessintl.com/august-2014/pdadmac-flocculant-reagent-use-clarisolve-depth-filters/* (Accessed: 2 April 2022).

Singh, N. et al. (2016) ‘Clarification technologies for monoclonal antibody manufacturing processes: Current state and future perspectives’, *Biotechnology and Bioengineering*, 113(4), pp. 698–716. doi:10.1002/bit.25810.

Siperstein, M. and Fagan, V. (1964) ‘Deletion of the Cholesterol-negative Feedback System in Liver Tumors’, *Cancer Res.*, 24(7), pp. 1108–1115.

Soares, E. V. (2011) ‘Flocculation in Saccharomyces cerevisiae: a review’, *Journal of Applied Microbiology*, 110(1), pp. 1–18. doi.org/10.1111/j.1365-2672.2010.04897.x.

Sontag, J.M. and Sontag, E. (2006) ‘Regulation of cell adhesion by PP2A and SV40 small tumor antigen: An important link to cell transformation’, *Cell. Mol. Life Sci.* 63, pp. 2979–2991. doi.org/10.1007/s00018-006-6300-7.

Srinivasan, S. et al. (2003) ‘Glucose Regulates Monocyte Adhesion Through Endothelial Production of Interleukin-8’, *Circulation Research*, 92(4), pp. 371–377. doi:10.1161/01.RES.0000061714.74668.5C.

Stahley, S.N. et al. (2014) 'Desmosome Assembly and Disassembly Are Membrane Raft-Dependent', *PLOS ONE*, 9(1), p. e87809. doi:10.1371/journal.pone.0087809.

Stanton, Dan (2021) 'Global mammalian capacity to reach 7,500kL by 2025 - Bioprocess Insider', BioProcess International, 4 November. Available at: *https://bioprocessintl.com/bioprocess-insider/facilities-capacity/spurred-by-cdmos-global-mammalian-capacity-to-reach-7500kl-by-2025/* (Accessed: 6 March 2022).

Stewart, G.G. and Goring, T.E. (1976) 'Effect of Some Monovalent and Divalent Metal Ions on the Flocculation of Brewers Yeast Strains', *Journal of the Institute of Brewing*, 82(6), pp. 341–342. doi:10.1002/j.2050-0416.1975.tb06959.x.

Stewart, G.G. (2018) 'Yeast Flocculation—Sedimentation and Flotation', *Fermentation*, 4(2), p. 28. doi.org/10.3390/fermentation4020028.

Strasser, L. et al. (2021) 'Proteomic Profiling of IgG1 Producing CHO Cells Using LC/LC-SPS-MS3: The Effects of Bioprocessing Conditions on Productivity and Product Quality', *Frontiers in Bioengineering and Biotechnology*, 9, p. 270. doi:10.3389/fbioe.2021.569045.

Stratford, M. (1989) 'Yeast flocculation: Calcium specificity', *Yeast*, 5(6), pp. 487–496. doi.org/10.1002/yea.320050608.

Stratford, M. (1992) 'Lectin-mediated Aggregation of Yeasts — Yeast Flocculation', *Biotechnology and Genetic Engineering Reviews*, 10(1), pp. 283–342. doi.org/10.1080/02648725.1992.10647891.

Takeichi, M. and Okada, T.S. (1972) 'Roles of magnesium and calcium ions in cell-to-substrate adhesion', *Experimental Cell Research*, 74(1), pp. 51–60. doi:10.1016/0014-4827(72)90480-6.

Talbot, N.J. et al. (1996) 'MPG1 Encodes a Fungal Hydrophobin Involved in Surface Interactions during Infection-Related Development of Magnaporthe grisea', *The Plant Cell,* Vol. 8. pp. 985-999.

Tatham, M.H. et al. (2011) 'Comparative proteomic analysis identifies a role for SUMO in protein quality control', *Science Signaling*, 4(178), p. rs4. doi:10.1126/scisignal.2001484.

Taylor, N.W. and Orton, W.L. (1973) 'Effect of Alkaline-Earth Metal Salts on Flocculence in Saccharomyces Cerevisiae', *Journal of the Institute of Brewing*, 79(4), pp. 294–297. doi:10.1002/j.2050-0416.1973.tb03543.x.

Tomic, S. et al. (2015) 'Complete clarification solution for processing high density cell culture harvests', *Separation and Purification Technology*, 141(5), pp. 269–275. doi.org/10.1016/j.seppur.2014.12.002.
Tvaroška, I., Selvaraj, C. and Koča, J. (2020) 'Selectins—The Two Dr. Jekyll and Mr. Hyde Faces of Adhesion Molecules—A Review', *Molecules*, 25(12), p. 2835. doi:10.3390/molecules25122835.
Ulgen, E., Ozisik, O. and Sezerman, O.U. (2019) 'pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks', *Frontiers in Genetics*, 10. doi:10.3389/fgene.2019.00858.
Urbich, C. et al. (2002) 'Shear Stress–Induced Endothelial Cell Migration Involves Integrin Signaling Via the Fibronectin Receptor Subunits α5 and β1 | Arteriosclerosis, Thrombosis, and Vascular Biology', *Arteriosclerosis, Thrombosis and Vascular Biology*, 22(1), pp. 69–75. doi:10.1161/hq0102.101518.
Vasaikar, S.V. et al (2021) 'EMTome: a resource for pan cancer analysis of epithelial-mesenchymal transition genes and signatures.' *Br. J. Cancer,* 124, pp259-269*.* doi:10.1038/s41416-020-01178-9
Valente, K.N. et al. (2018) 'Applications of proteomic methods for CHO host cell protein characterization in biopharmaceutical manufacturing', *Current Opinion in Biotechnology*, 53, pp. 144–150. doi:10.1016/j.copbio.2018.01.004.
Van Campenhout, C.A. et al. (2011) 'Dlg3 Trafficking and Apical Tight Junction Formation Is Regulated by Nedd4 and Nedd4-2 E3 Ubiquitin Ligases', *Developmental Cell*, 21(3), pp. 479–491. doi.org/10.1016/j.devcel.2011.08.003.
Veelders, M. et al. (2010) 'Structural basis of flocculin-mediated social behavior in yeast', *Proceedings of the National Academy of Sciences*, 107(52), pp. 22511–22516. doi:10.1073/pnas.1013210108.
Waheed, A.A. and Jones, T.L.Z. (2002) 'Hsp90 Interactions and Acylation Target the G Protein Gα12 but Not Gα13 to Lipid Rafts', *Journal of Biological Chemistry*, 277(36), pp. 32409–32412. doi.org/10.1074/jbc.C200383200.
Walling, C. and Weil, T. (1974) 'The ferric ion catalyzed decomposition of hydrogen peroxide in perchloric acid solution', *International Journal of Chemical Kinetics*, 6(4), pp. 507–516. doi:10.1002/kin.550060406.
Walther, C.G., Whitfield, R. and James, D.C. (2016) 'Importance of Interaction between Integrin and Actin Cytoskeleton in Suspension Adaptation of CHO cells',

Applied Biochemistry and Biotechnology, 178(7), pp. 1286–1302. doi:10.1007/s12010-015-1945-z.

Wang, Y. et al. (2007) 'Human 1A6/DRIM, the homolog of yeast Utp20, functions in the 18S rRNA processing', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1773(6), pp. 863–868. doi:10.1016/j.bbamcr.2007.04.002.

Wasinger, V.C. (1995) 'Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium', *Electrophoresis*, 16, pp. 1090–1094. doi:doi:10.1022/elps.11501601185.

Weber-Boyvat, M. et al. (2015) 'OSBP-related protein 3 (ORP3) coupling with VAMP-associated protein A regulates R-Ras activity', *Experimental Cell Research*, 331, pp. 278–291. doi: 10.1016/j.yexcr.2014.10.019.

Wessels, J.G.H. (1991a) 'The thn mutation of Schizophyllum commune, which suppresses formation of aerial hyphae, affects expression of the Sc3 hydrophobin gene', J. Gen. Micro, 137, pp. 2439–2445.

Wessels, J.G.H. (1991b) 'Hydrophobin Genes lnvolved in Formation of Aerial Hyphae and Fruit Bodies in Schizophyllum', *The American Society of Plant Physiologists*, 3, pp. 793–799.

Wessels, J. (2000) 'Hydrophobins, unique fungal proteins', *Mycologist*, 14(4), pp. 153–59.

Westoby, M. et al. (2011) 'Effects of solution environment on mammalian cell fermentation broth properties: Enhanced impurity removal and clarification performance', *Biotechnology and Bioengineering*, 108(1), pp. 50–58. doi.org/10.1002/bit.22923.

Wolf, A. et al. (2013) 'Insulin signalling via Akt2 switches plakophilin 1 function from stabilizing cell adhesion to promoting cell proliferation', *Journal of Cell Science*, 126(8), pp. 1832–1844. doi:10.1242/jcs.118992.

Wong, W.C., Dye, D.E. and Coombe, D.R. (2012) 'The Role of Immunoglobulin Superfamily Cell Adhesion Molecules in Cancer Metastasis', *International Journal of Cell Biology*, 2012, p. e340296. doi:10.1155/2012/340296.

Wösten, H.A.B. (1993) 'lnterfacial Self-Assembly of a Funga1 Hydrophobin into a Hydrophobic Rodlet Layer', *The Plant Cell*, 5, pp. 1567–1574.

Wösten, H.A.B. (2001) 'Hydrophobins: Multipurpose Proteins', *Annual Review of Microbiology*, 55(1), pp. 625–646. doi:10.1146/annurev.micro.55.1.625.

Wurm, M.J. and Wurm, F.M. (2021) 'Naming CHO cells for bio-manufacturing: Genome plasticity and variant phenotypes of cell populations in bioreactors question the relevance of old names', *Biotechnology Journal*, 16(7), p. 2100165. doi:10.1002/biot.202100165.

Xie, K. and Huang, S. (2003) 'Regulation of cancer metastasis by stress pathways', *Clinical & Experimental Metastasis*, 20(1), pp. 31–43. doi:10.1023/a:1022590402748.

Xiong, Q. et al. (2015) 'Proteomic study of different culture medium serum volume fractions on RANKL-dependent RAW264.7 cells differentiating into osteoclasts', *Proteome Science*, 13(1), p. 16. doi:10.1186/s12953-015-0073-6.

Xu, F. et al. (2005) 'Dual roles for cholesterol in mammalian cells', *Proceedings of the National Academy of Sciences*, 102(41), pp. 14551–14556. doi:10.1073/pnas.0503590102.

Xu, N. et al. (2017) 'Comparative Proteomic Analysis of Three Chinese Hamster Ovary (CHO) Host Cells', *Biochemical engineering journal*, 124, pp. 122–129. doi:10.1016/j.bej.2017.05.007.

Xu S. et al. (2013), 'CD36 enhances fatty acid uptake by increasing the rate of intracellular esterification but not transport across the plasma membrane', *Biochemistry*, 52(41), pp.7254-61. doi: 10.1021/bi400914c.

Xue, J. et al. (2020) 'Enhanced in-Source Fragmentation Annotation Enables Novel Data Independent Acquisition and Autonomous METLIN Molecular Identification', *Analytical Chemistry*, 92(8), pp. 6051–6059. doi:10.1021/acs.analchem.0c00409.

Xu, X. et al. (2011) 'The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line', *Nature Biotechnology*, 29(8), pp. 735–741. doi.org/10.1038/nbt.1932.

Yang, Shucai et al. (2017) 'FOXP3 promotes tumor growth and metastasis by activating Wnt/β-catenin signaling pathway and EMT in non-small cell lung cancer', *Molecular Cancer*, 16(1), p. 124. doi:10.1186/s12943-017-0700-1.

Yanai, H. et al. (2016) 'Tissue repair genes : the TiRe database and its implications for skin wound healing.' *Oncotarget*, 7(16), p 21145-21155. doi:10.18632/oncotarget.8501.

Yao, T. and Asayama, Y. (2017) 'Animal-cell culture media: History, characteristics, and current issues'*, Reproductive Medicine and Biology*, 16(2), pp. 99–117. doi:10.1002/rmb2.12024.

Yoo, Y.A. et al. (2011) ‘Sonic hedgehog pathway promotes metastasis and lymphangiogenesis via activation of Akt, EMT, and MMP-9 pathway in gastric cancer.’, *Cancer research* [Preprint]. doi:10.1158/0008-5472.CAN-11-1338.

Yu, G. et al. (2012) ‘clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters’, *OMICS : a Journal of Integrative Biology*, 16(5), pp. 284–287. doi.org/10.1089/omi.2011.0118.

Yu, H. et al. (2018) ‘Inhibition of cell migration by focal adhesion kinase: Time-dependent difference in integrin-induced signaling between endothelial and hepatoblastoma cells’, *International Journal of Molecular Medicine*, 41(5), pp. 2573–2588. doi.org/10.3892/ijmm.2018.3512.

Yue, B. (2014) ‘Biology of the Extracellular Matrix: An Overview’, *Journal of glaucoma*, pp. S20–S23. doi:10.1097/IJG.0000000000000108.

Zaniolo, K. et al. (2006) ‘Expression of the Gene Encoding Poly(ADP-ribose) Polymerase-1 Is Modulated by Fibronectin during Corneal Wound Healing’, *Investigative Ophthalmology & Visual Science*, 47(10), pp. 4199–4210. doi:10.1167/iovs.06-0176.

Zhang, F.L. and Casey, P.J. (1996) ‘PROTEIN PRENYLATION: Molecular Mechanisms and Functional Consequences’, *Annual Review of Biochemistry*, 65(1), pp. 241–269. doi:10.1146/annurev.bi.65.070196.001325.

Zhang, H. et al. (2018) ‘Regulatory Mechanisms of the Wnt/β-Catenin Pathway in Diabetic Cutaneous Ulcers’, *Frontiers in Pharmacology*, 9, p. 1114. doi:10.3389/fphar.2018.01114.

Zhang, Y. et al. (2020) ‘Activation of PGK1 under hypoxic conditions promotes glycolysis and increases stem cell-like properties and the epithelial-mesenchymal transition in oral squamous cell carcinoma cells via the AKT signalling pathway’*, International Journal of Oncology*, 57(3), pp. 743–755. doi:10.3892/ijo.2020.5083.

Zhao, X. and Guan, J.-L. (2011) ‘Focal adhesion kinase and its signalling pathways in cell migration and angiogenesis’, *Advanced Drug Delivery Reviews*, 63(8), pp. 610–615. doi:10.1016/j.addr.2010.11.001.

Zhong, X. et al. (2015) ‘Human cell adhesion molecules: annotated functional subtypes and overrepresentation of addiction-associated genes: Cell adhesion molecules and dopaminergic connections’, *Annals of the New York Academy of Sciences*, 1349(1), pp. 83–95. doi:10.1111/nyas.12776.

Zhu, M.M. et al. (2008) ‘Effects of Elevated pCO2 and Osmolality on Growth of CHO Cells and Production of Antibody-Fusion Protein B1: A Case Study’, *Biotechnology Progress*, 21(1), pp. 70–77. doi:10.1021/bp049815s.

APPENDICES

List of Appendices

The appendices for this thesis are available electronically upon request due to the size/amount of data in them.

Supplementary Material for Chapter 3

Appendix 1 This appendix contains the GeneBank gene sequence for all five proteins.

Appendix 2 This appendix contains the results of the Protter analysis for expected subcellular expression location.

Supplementary Material for Chapter 5

Appendix 3 All DE proteins detected in the A5 v S5 comparison for CHO K1 cells.

Appendix 4 Enriched KEGG pathways from the A5 v S5 comparison for CHO K1 cells.

Appendix 5 All DE proteins detected in the A0 v S0 comparison for CHO K1 cells.

Appendix 6 Enriched KEGG pathways from the A0 v S0 comparison for CHO K1 cells.

Appendix 7 Enriched GO terms from the A5 v S5 comparison for CHO K1 cells.

Appendix 8 Enriched GO terms from the A0 v S0 comparison for CHO K1 cells.

Supplementary Material for Chapter 6

Appendix 9 All DE proteins detected in CHO K1 cells following adaptation to suspension growth.

Appendix 10 All DE proteins detected in CHO DXB11 cells following adaptation to suspension growth.

Appendix 11 DE proteins common to both CH K1 and CHO DXB11 cells following adaptation to suspension growth.

Appendix 12 Enriched KEGG pathways based on the DE proteins from appendix 11.

Appendix 13 Enriched GO terms based on the DE proteins from appendix 11.

Appendix 14 List of published proteins associated with the epithelial mesenchymal transition (EMT).

Appendix 15 List of published proteins associated with wound healing.

Appendix 16 List of KEGG pathways related to cellular adhesion.

Appendix 17 List of GO terms related to cellular adhesion.

Appendix 18 DE proteins detected in CHO K1 cells following adaptation to suspension growth (1% FDR Filter applied).

Appendix 19 All DE proteins detected in CHO DXB11 cells following adaptation to suspension growth (1% DFR filter applied).

Supplementary Material for Chapter 7

Appendix 20 All DE proteins detected in CHO K1 cells following adaptation to suspension growth in serum free media.

Appendix 21 All DE proteins detected in CHO DXB11 cells following adaptation to suspension growth in serum free media.

Appendix 22 DE proteins common to both CH K1 and CHO DXB11 cells following adaptation to suspension growth in serum free media.

Appendix 23 Enriched KEGG pathways based on the DE proteins from appendix 22.

Appendix 24 Enriched GO terms based on the DE proteins from appendix 22.

Appendix 25 DE proteins detected in CHO K1 cells following adaptation to suspension growth in serum free media (1% FDR Filter applied).

Appendix 26 All DE proteins detected in CHO DXB11 cells following adaptation to suspension growth in serum free media (1% DFR filter applied).

Appendix 27 Enriched GO terms based on the thirty DE proteins common to appendices 25 and 26.