High throughput DNA single cell analysis of CHO-K1 cell surface glycosylation using lectin probes

Thesis submitted for the award of

Doctor of Philosophy

by

Flávio Ferreira, B.Eng., PGDip

Supervised by

Brendan O'Connor, B.Sc., Ph.D

and

Dermot Walls, B.Sc., Ph.D

School of Biotechnology

Dublin City University

December 2019

Declaration

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

Signed: _____ ID No.: 12211732

Date: _____

Acknowledgments

Firstly, I would like to thank my family, particularly my mom and sister for their invaluable support throughout my academic life.

I want to thank all my friends who I have had the pleasure to share the achievements and challenges which come with a PhD life.

In addition, I want to take the opportunity to express my gratitude to Brendan O'Connor and Dermot Walls who trusted me the development of this PhD work and for their continuous encouragement.

Table of contents

Abbreviations

7-AAD	7-Aminoactinomycin D
AAL	*Aleuria aurantia* lectin
AAL-2	*Aleuria aurantia* lectin 2
ACS	American Chemical Society
ADC	Analog-to-Digital Converter
APS	Ammonium persulphate
ASAL	*Allium sativum* leaf agglutinin
Asn	Asparagine
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
BSC	Biological Safety Cabinet
CHO	*Chinese Hamsters Ovary*
CON A	Concanavalin A
CS&T	Characterization, setup and tracking
CV	Column volume
Da	dalton
DF	Dilution factor
df	Degrees of freedom
dH2O	distilled H2O
DHFR	Dihydrofolate reductase
DMSO	Dimethyl sulfoxide
DRAQ5	Deep red-fluorescing bisalkylaminoanthraquinone number five

DTT	Dithiothreitol
ELLA	Enzyme-linked lectin assay
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated protein degradation
FACS	Fluorescence-activated cell sorting
FITC	Fluorescein isothiocyanate
FSC	Forward-scattered light
Fuc	Fucose
g	gram
Gal	Galactose
GalNac	*N*-acetylgalactosamine
GE	General Electric
Glc	glucose
GlcNAc	*N*-acetylglucosamine
GlcNAc –TV	*N*-acetylglucosomyltransferase V
Glu	Glucose
GRFT	Griffithsin lectin
GST	glutathione S-transferase
H	Height
HRP	Horseradish peroxidase
IEC	Intestinal epithelial cells
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IMAC	Immobilised Metal Affinity Chromatography
IMS	Industrial Methylated Spirits

IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Lysine
kDa	kilodalton
kpsi	kilopound per square inch
L	Liter
L-PHA	Leucoagglutinin *Phaseolus vulgaris* agglutinin
Lac	Lactose
LB	Luria Bertani
LEC A	Lectin A
LEC B	Lectin B
M	molar
mA	milliampere
mAB	Monoclonal antibody
MAH	*Maackia amurensis*
MAL II	*Maackia Amurensis* Lectin II
Man	mannose
mg	miligram
MIR	Mid infrared
mL	mililiter
mM	milimol
mm	milimeter
MS	Mass spectrometry
MWCO	molecular weight cut-off
MX	α-mannosidase-IIx
NaBu	Sodium butyrate

NHS	*N*-Hydroxysuccinimide
NICB	National Institute for Cellular Biotechnology
NIR	Near infrared
nm	nanometer
O-GlcNAc	*O*-linked β-*N*-acetylglucosamine
PBS	Phosphate-buffered saline
pCO2	Partial pressure of carbon dioxide
PE	Phycoerythrin
PEG	Polyethylene glycol
PI	Propidium iodide
PNA	Peanut agglutinin
PTM	Photomultiplier tube
PTM	Post-translational modification
PVA	Polyvinyl alcohol
RCA I	*Ricinus communis* Agglutinin I
rpm	Rotation per minute
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SDTB	Semi Dry Transfer Buffer
Ser	Serine
Sia	Sialic acid
SSC	Side-scattered light
TB	Terrific Broth
TBST	Tris-Buffered Saline and Tween 20
TEMED	*N,N,N',N'*-tetramethylethane-1,2-diamine

TFS	Thermo Fisher Scientific
Thr	Threonine
TMB	3,3’,5,5’-Tetramethylbenzidine
tPA	tissue plasminogen activator
UPR	Unfolded protein response
V	volts
v/v	Volume per volume
W	Width
w/v	Weight per volume
WGA	Wheat germ agglutinin
WR	Working reagent
µL	microliter

List of Figures

List of Tables

Abstract

High Throughput DNA Singe Cell Analysis of CHO-K1 cell surface glycosylation using lectin probes

Flávio Ferreira

Biological glycosylation is the process which adds specific sugars to other sugars, proteins and lipids. Protein glycosylation is one of the most important post-translational modifications, which occurs in more than half of all proteins present in the human body. Abnormal glycosylation has been demonstrated to be linked to many different diseases due to alterations associated with protein folding and biological function. Therefore, glycosylation is absolutely essential for the correct structure, function and stability of important proteins.

Surface glycosylation patterns play a key role in the modulation of the immune responses which are mediated by carbohydrate-binding proteins called Lectins. Such biomolecules are typically highly selective for specific glycan structures, making them extremely useful for glycan variation investigation.

A rapid and accurate bioanalytical method to detect early unhealthy cell signs during a bioprocess is a current issue facing the industry. It is widely known that as cells become stressed or diseased the earliest changes that occur are in cell surface glycosylation.

CHO cells are the host cell of choice of the rapidly emerging biopharmaceutical industry for the production of glycoprotein therapeutics. Hence, this research work investigated the interaction between lectin probes with the membrane glycoconjugates of CHO cells subjected to different levels of spent medium, temperature and CO_2.

High throughput DNA single cell analysis using flow cytometry allowed the determination of cell surface glycosylation variation in response to the stressors. Cells subjected to different levels of spent medium had their cell surface glycosylation profile most affected in relation to cells subjected to temperature and CO_2 alteration. Fucose and N-Acetylglucosamine were identified as key glycans changing on the cell surface.

1 Introduction

1.1 Glycobiology

Glycobiology is the study of the multiple functions of sugars, that is, carbohydrates attached to lipids and proteins. Carbohydrates are very complex and not encoded in the genome. This fact might have discouraged investigators from looking at the biological functions of sugar groups beyond the context of metabolism in cells. As a result, the study of nucleic acids, proteins and lipids has been the subject of great attention in the scientific field for over a century and only recently have carbohydrates received increased attention as the expansion of the field of glycobiology takes place (Ghazarian, Idoni and Oppenheimer, 2011).

Although the tendency is to assume that the biological information flows through only three classes of biomolecules, i.e., from DNA to RNA to protein, the enormous biological complexities found in the human body and many other organisms rely on two other major classes of biomolecules: lipids and carbohydrates. These molecules play a role in mediating the generation of energy and signalling responses to a stimulus. Also, they can act as recognition markers and structural components (Varki and Sharon, 2009). In the case of carbohydrates, the addition of sugar groups to proteins, a process which is known as glycosylation, encompasses one of the most crucial posttranslational modifications (PTMs) of proteins. Additionally, biological glycosylation is not a process which adds specific sugars to proteins, but also to lipids and to other sugar groups (glycans) (Marth and Grewal, 2008). Therefore, glycosylation helps to explain the reason why the relatively small number of genes in the genome is able to create the highly biological complexities found in organisms in relation to their development, growth, and functioning (Varki and Sharon, 2009).

Protein glycosylation occurs in more than half of all proteins present in the human body. Abnormal glycosylation has been demonstrated to be linked to many different diseases due to alterations associated with protein folding and biological function (Christiansen *et al.*,

2014). Therefore, glycosylation is absolutely essential for the correct structure, function and stability of important proteins (Lepenies and Seeberger, 2014).

Furthermore, glycosylation patterns on the surface of the cell plays a key role in the modulation of the immune responses (Veiseh *et al.*, 2014). These responses are *mediated by carbohydrate-binding proteins called Lectins* (Gorelik, Galili and Raz, 2001). Such biomolecules are typically highly selective for specific glycan structures making them (i) extremely useful for glycan variation investigation and (ii) perhaps the most largely studied biomolecules in the field of glycobiology (Ohtsubo and Marth, 2006).

1.1.1 Carbohydrate structure and function

Among the four major classes of organic molecules found in living organisms; proteins, lipids, nucleic acids and carbohydrates, the last are the most prevalent organic molecules present in nature (Wade, 1999). The general empirical formula of most simple sugars is $C_nH_{2n}O_n$, where n is $n \geq 3$. The proportion found between these three atoms suggests that carbon atoms are somehow combined with water molecules. For this reason, the term carbohydrate has been used to refer to organic molecules which fall within this empirical chemical formula (Wade, 1999). For instance, glucose is a very common monosaccharide which is broken down into carbon dioxide and water molecules through oxidation. The energy released from this reaction is used in cellular processes to carry out protein synthesis, movement and transport to name a few. In plant and animal systems, glucose molecules are combined to generate large molecules for energy storage such as starch and glycogen. On the other hand, glucose can actually be combined in different ways to create an assortment of other macromolecules. Cellulose is a glucose based macromolecule which is found in the cell wall of plants. Glucose molecules are linked through β-1,4 glycosidic bonds to form cellulose whereas in starch, glucose monomers are combined through α-1,4 glycosidic bonds, and in glycogen the combination is through α-1,4 and α-1,6 glycosidic bonds (Wade, 1999).

Carbohydrates in living organisms are extremely heterogeneous due to a variety of features found in this class of organic molecules (see Figure 1.1). Different types and numbers of glycan residues have the ability to link together to create glycosidic bonds with each other. The carbohydrate structural characteristics, the anomeric linkage type, the location and the absence or presence of branching are the main features providing sugar molecules with a high level of complexity in heterogeneity (Mody, Joshi and Chaney, 1995; Gorelik, Galili and Raz, 2001). In order to grasp the level of complexity of sugar molecules, one may look at the comparison between a single disaccharide molecule and a single dipeptide one. These two molecules are composed of two identical molecules such as a single hexose monosaccharide for the former and a single amino acid such as glycine for the latter. Two hexose monosaccharides can form 11 different disaccharides whereas two glycine molecules can only form a single dipeptide. A considerable escalation in the level of complexity of carbohydrates can be observed when looking at the comparison between the number of potential unique combinations that four different amino acid molecules and four different hexose monosaccharide molecules can produce. While these 4 amino acids may produce 24 different tetrapeptides, the monosaccharides may possibly form 35,560 varieties of tetrasaccharide molecules (Sharon and Lis, 1989, 1993).

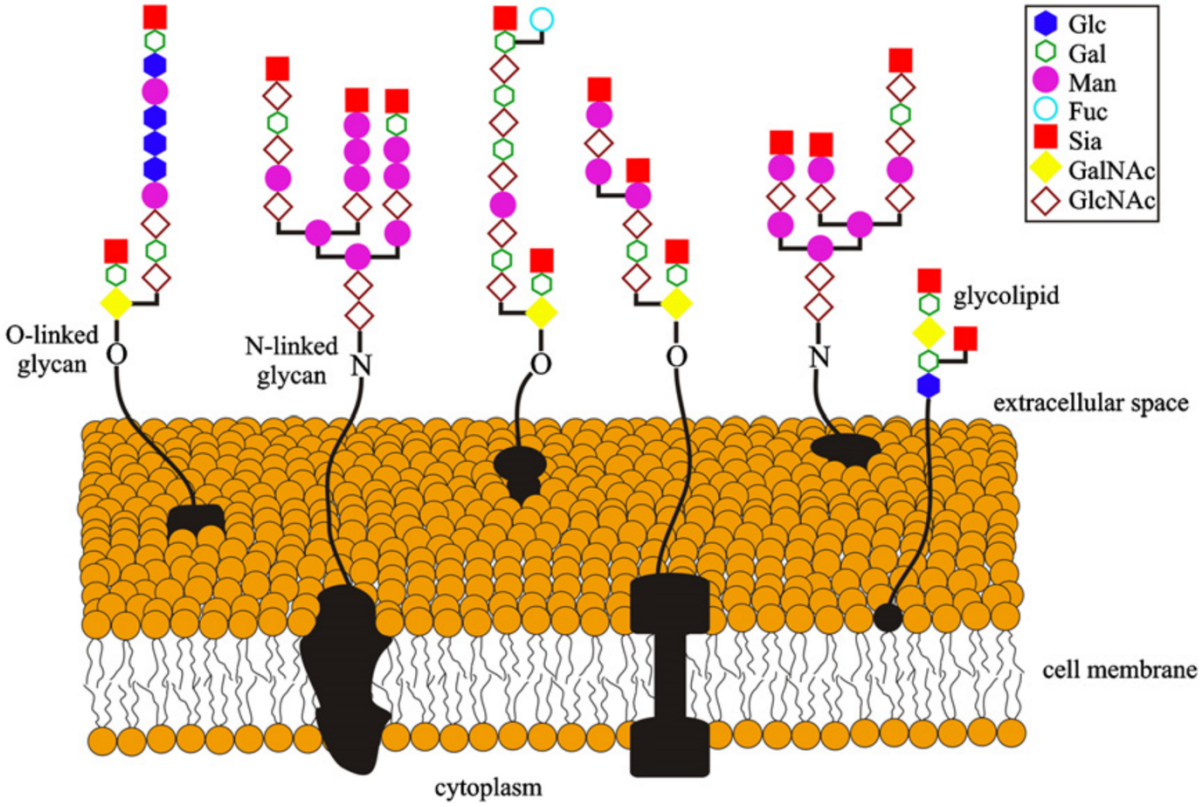

Figure 1.1: Schematic diagram illustrating the heterogeneity of sugar structures found on cell surface glycoconjugates (glycoproteins and glycolipids). *O*-linked and *N*-linked glycans of glycoconjugates are normally terminated with Sialic acid (Sia) residues. Glc = glucose; Gal = galactose; Man = mannose; Fuc = fucose; GalNac = *N*-acetylgalactosamine; GlcNAc = *N*-acetylglucosamine. Figure was extracted from Ghazarian et al. 2011.

Although sugar molecules present a large diversity in biological information, such molecules are not encoded by the genome (Feizi & Mulloy 2003). However, the genome encodes enzymes such as glycosyltransferases and glycosidases which act upon glycans. On different activity levels, these enzymes work together in the endoplasmic reticulum (ER) and the Golgi apparatus, defining the patterns of glycosylation of glycoconjugates (glycoproteins and glycolipids) (Ghazarian, Idoni and Oppenheimer, 2011).

The complexity and structural variability of glycosylation patterns on cell surface sugars, that is, glycan structures attached to proteins and lipids on the cell wall, provides these glycans with the ability to perform signalling, recognition and adhesion functions (Ofek, Hasty and

Sharon, 2003; Varki and Sharon, 2009). Therefore, the sugars on the cell surface are involved in several physiological functions of great importance such as oogenesis, spermatogenesis and normal embryonic development, differentiation, growth, contact inhibition, cell-cell binding and recognition, cell signalling, host-pathogen interactions in infectious process, immune response of a host cell, the development of diseases, metastasis, intracellular trafficking and localization, the rate of degradation and membrane rigidity (Subtelny and Wessells, 1980; Sharon and Lis, 1993; Kennedy *et al.*, 1995; Ghazarian, Idoni and Oppenheimer, 2011).

1.1.2 *N*-linked and *O*-linked oligosaccharides

Large carbohydrate molecules, that is, oligosaccharides, can be linked to proteins through glycosidic bonds by two types of linkages: *N*-linked and *O*-linked. The first one consists of the binding of *N*-acetylglucosamine to the amide side chain of asparagine. Asn-X-Ser(Thr)- is the sequence of asparagine residues found in *N*-linked oligosaccharides. The X position can be of any amino acid except to proline (Gorelik, Galili and Raz, 2001). The second type of linkage , the *O*-linked, consists of the binding of C-1 of *N*-acetylgalactosamine to the hydroxyl of serine or threorine amino acids (Gorelik, Galili and Raz, 2001).

The glycosylation of *N*-linked oligosaccharides in eukaryotes starts with the covalent binding of a 14 long common oligosaccharide precursor composed of 2 *N*-acetylglucosamine, 9 mannose and 3 glucose molecules to the asparagine residue of the newly synthesized target protein chain as this protein is transported into the endoplasmic reticulum (ER). As a result of the addition of this common oligosaccharide precursor to the polypeptide, fully processed *N*-linked carbohydrates can be classified into three major classes: high-mannose, complex and hybrid oligosaccharides. In order for some eukaryotic proteins to be properly folded, *N*-linked glycosylation must be carried out. As these proteins are correctly folded, three glucose residues are removed from the 14 long oligosaccharide and the proteins are transported from the ER to the Golgi apparatus. Depending on how the oligosaccharide is processed in the Golgi apparatus, carbohydrates are then classified into the three aforementioned classes. The oligosaccharide which does not suffer any removal or addition of monosaccharides is classified

as high-mannose oligosaccharides. Oligosaccharide which may have mannose residues removed or may have other monosaccharides added, falls in the complex or hybrid class (Ghazarian, Idoni and Oppenheimer, 2011). High mannose and complex oligosaccharides share a common core structure but differ in the terminal elaborations which extend from this common core (Taylor and Drickamer, 2006).

The *O*-linked glycosylation is an alteration of glycoproteins which is highly likely to take place in the Golgi apparatus (Röttger *et al.*, 1998; Patsos *et al.*, 2009). In *O*-linked carbohydrates the C-1 of *N*-acetygalactosamine is covalently linked to the hydroxyl of threonine or serine of the protein chain (Röttger *et al.*, 1998; Patsos *et al.*, 2009). After the addition of the *N*-acetylgalactosamine residue to the protein chain, the oligosaccharide may be extended by the addition of monosaccharides (e.g. galactose, fucose, *N*-acetylglucosamine and sialic acid) (Schachter and Brockhausen, 1992; Mitra *et al.*, 2006). Many unique *O*-linked carbohydrates have been investigated such as *O*-fucose, *O*-mannose and *O*-*N*-acetylglucosamine. Studies have demonstrated that the modification process of proteins with *O*-linked β-*N*-acetylglucosamine (*O*-GlcNAc) has an effect on the protein biological function via a number of mechanisms; for example, protein function alteration due to phosphorylation, protein-protein interaction regulation, protein degradation regulation, protein localization and transcription regulation (Hanover, 2001; Zachara and Hart, 2006; Hart, Housley and Slawson, 2007).

1.1.3 Modifications of cell surface carbohydrates

Numerous studies have demonstrated alterations in cell surface glycosylation as cells go through different stages in the biological development such as differentiation and embryonic development (Balcan *et al.*, 2008; Park *et al.*, 2015; Delannoy *et al.*, 2017). Additionally, modifications in the cell surface glycosylation profile in inflammatory and cancerous processes have also been extensively reported by several scientific studies, indicating carbohydrates as potential biomarkers for the identification of the onset of diseases (Veiseh et al. 2014; Krasnewich 2014; Gorelik et al. 2001; Patsos et al. 2009; An et al. 2009).

The bio-production of sugars in a cell relies on several highly competitive processes which involves glycosyltransferases and enzymes responsible for catalysing the formation of the glycosidic linkage. This fact makes the glycosylation process greatly sensitive to the biochemical environment and alterations in glycosylation patterns can potentially implicate in many diseases such as cancer, gastrointestinal related diseases and cognitive impairments (An et al. 2009; Krasnewich 2014). The functionality of glycoproteins and glycolipids is affected by the glycosylation process, as this process modifies the physical properties of glycoconjugates. The protein folding process is highly influenced by the specific sugars attached to particular sites of the protein and some of these sugars on glycoconjugates can also play an important role on the process of specific sugar recognition by glycan-binding proteins.

A study in biological development showed the great importance of glycosylation in spermatogenesis. *Man2a2* is a gene which encodes α-mannosidase-IIx (MX), an enzyme associated with the synthesis of *N*-glycan intermediates. This gene was disrupted and male MX-null mice developed small testes and were infertile (Akama *et al.*, 2002). It was observed that germ cells failed to adhere to Sertoli cells in the seminiferous tubules and, as a result, the developing germ cells were prematurely released from the seminiferous epithelium to the epididymis. MX enzyme in germ cells is associated with the biosynthesis of a GlcNAc-terminated triantennary and fucosylated *N*-glycan structure. Such oligosaccharide on the surface of the cell may play a critical role in the adhesion process between germ cells and Sertoli cells (Akama *et al.*, 2002).

Glycosylation changes on cell surface of intestinal epithelial cells (IEC) were correlated with glycosyltransferase activities during cell differentiation process (Park *et al.*, 2015). As the cells differentiated, a decrease in high mannose type glycans was observed and also a simultaneous increase in fucosylated and sialyated complex/hybrid carbohydrates. An increase in activity was observed for GlcNAc transferase II and V, which are enzymes involved in *N*-glycosylation (Brockhausen, Romero and Herscovics, 1991). Also, β-3-galactosyltransferase, α-2-fucosyltransferase, sialyltransferase, and β-6-GlcNAc transferase, which are enzymes critical to *O*-glycan biosyntheses, were all increased in activity (Amano, Kobayashi and Oshima, 2001).

At day 21 of the study, when cells seemed to be fully differentiated, the changes in glycosylation on the cell surface terminated (Park *et al.*, 2015).

The investigation of alterations of cell surface glycosylation in tumors from cancer patients and experimental animal demonstrated that the majority of the modifications observed in membrane glycoproteins involved the presence of larger, more branched *N*-linked carbohydrates, more specifically, β1-6GlcNAc-branched *N*-linked glycans (Dennis, 1991, 1992; Fernandes *et al.*, 1991). The increase of these oligosaccharides in the cells was detected by observing the increase in L-PHA (leucoagglutinin *Phaseolus vulgaris* agglutinin) lectin interaction (Dennis, 1991, 1992; Fernandes *et al.*, 1991). An increased activity of *N*-acetylglucosomyltransferase V (GlcNAc –TV) results in the increase of GlcNAcβ1-6Manα1-6Manβ branching at the trimannosyl core of complex-type carbohydrates which, in turn, increases the β1-6 branching of *N*-linked glycans. The increase of β1-6 branching of *N*-linked oligosaccharides was observed in the early stage of tumour development which was induced by oncogenes v-src, H-ras, v-fps or oncogenic virus (Yamashita *et al.*, 1985; Pierce and Arango, 1986; Dennis *et al.*, 1989).

1.1.4 Lectins

Lectins consist of a very heterogeneous group of proteins with specific capabilities to selectively recognize and reversibly bind to specific glycans on glycoconjugates without modifying the carbohydrate structures. Lectins not only bind to oligosaccharides on cells but to free sugars as well, including monosaccharides (Lannoo and Damme, 2010). Lectins are multivalent and also referred to as agglutinins due to the fact that the majority of them have cell agglutination capability. Lectins were first discovered in plants and for a long time they were believed to be present in plant organisms only. However, lectins were subsequently also found in different organisms such as bacteria, viruses, fungus and in humans. Although the presence of lectins is ubiquitous in living systems, plants contain lectins in the largest quantity; thus, plant lectins have been extensively scientifically investigated, particularly those sourced from legumes. Lectins in plants are mostly confined to seeds of legumes, roots, tubers, bulbs,

bark, leaves, tissues of flowers, and other tissues and organs (Harold and Gabius, 2001; De Mejía and Prisecaru, 2005). Foods such as wheat, corn, tomatoes, peanuts, kidney beans, bananas, peas, lentils, soybeans, mushrooms, rice and potatoes also contain lectins (De Mejía and Prisecaru, 2005). The biological role of lectins involves cell recognition, interaction and adhesion. In plants, lectins play a role in defense and symbiosis processes (Chrispeels and Raikhelb, 1991; Peumans and Van Damme, 1995; De Hoff, Brill and Hirsch, 2009; Michiels, Van Damme and Smagghe, 2010).

The classification of lectins was initially based on carbohydrate specificity and subsequently lectins were grouped according to subunit structures such as merolectins and hololectins. Another classification adopted was by families; for example, legume lectins and monocot mannose-binding lectins (Lam and Ng, 2011). Finally, plant lectins were classified into 12 different families according to three features: carbohydrate-binding domains, three-dimensional structures and the sequence of amino-acids (Van Damme, Lannoo and Peumans, 2008). Lectins are involved in many phenomena of biological recognition and these proteins have a multitude of different biological activities such as immunomodulatory, anti-insect, anti-viral, anti-tumor and anti-microbial (Jagtap and Bapat, 2010; Lam and Ng, 2011). For this reason, lectins have been largely applied in several areas; for example, biochemistry, cell biology and biomedicine. In the latter one, lectins have been used for the development of biomedical diagnostics tools and therapeutics (De Mejía and Prisecaru, 2005; Mislovičová *et al.*, 2009; Liu, Bian and Bao, 2010). Additionally, lectins have been used to develop drug delivery systems for specific anti-tumor therapy (Ghazarian, Idoni and Oppenheimer, 2011). The most significant recent advancement in the study of glycobiology is the introduction of lectins in microarrays or biosensors. This has allowed the examination of protein glycosylation and cell glycoprofiling in a high throughput manner (Rosenfeld *et al.*, 2007; Gemeiner *et al.*, 2009; Gupta, Surolia and Sampathkumar, 2010; Rahaie and Kazemi, 2010).

1.1.5 Recombinant lectins

A large variety of lectins from several organisms has been fully characterized. However, there are numerous disadvantages to obtain lectins from natural sources. For instance, the process is time-consuming and requires a considerable amount of biomass. With exception of lectins present in seeds and vegetative storage tissues, lectin yields are significantly low, purified lectins can contain undesired biomolecules, and a considerable ‘‘batch to batch’’ variation of the lectin source which can result in heterogeneity in the binding properties of lectins (Gemeiner *et al.*, 2009). Nevertheless, the expression and production of recombinant lectins in heterologous systems can overcome many of the issues encountered in the extraction of lectins from natural sources.

The production of recombinant lectins provides lectins with higher level of purity as well as a defined sequence of the amino acids involved. Therefore, the final features of the biomolecule are inevitably under more control. Furthermore, higher lectin yields can be achieved within a much shorter time (Gemeiner *et al.*, 2009).

Multiple plant lectins and lectins from different organisms have been expressed in bacteria (*Escherichia coli*), yeasts (*Pichia pastoris* and *Saccharomyces cerevisiae*), cells of insects (*Spodoptera frugiperda* ovarian cells) and mammalian cells such as monkey kidney cells (Oliveira, Texeira and Domingues, 2013). As *Escherichia coli* is the expression system of choice for this present research work, it is the intention to present in the following section a full analysis of the factors concerning the production of recombinant lectins in this bacterium.

1.1.6 Production of recombinant lectins in *Escherichia coli*

Escherichia coli expression system has several advantages such as rapid growth and expression rate, simple genome manipulations and cultivation, low cost and time, high yields can be obtained, the system can be scaled up and it is suitable for the production of lectins

which do not require post-translational modifications (Yin *et al.*, 2007; Demain and Vaishnav, 2009).

In addition, *E. coli* is the first expression system to hold the cDNA during the cloning process in the majority of the cases (Streicher and Sharon, 2003). The preferred vehicle used for lectin expression in *E. coli* is the pET expression vector family which is available in more than 40 configurations (Sørensen and Mortensen, 2005; Yin *et al.*, 2007). High expression levels of recombinant lectins can be achieved as the strong lac promoter (T7 promoter) which is present in pET vectors is induced with isopropyl-β-d-thiogalactopyranoside (IPTG) (Oliveira, Texeira and Domingues, 2013).

Recombinant lectins are frequently expressed as fusion proteins in order to ease the purification process. The expressed lectin has its cDNA ligated to a fusion partner which maintains the correct reading frame. Fusion partners are composed of a peptide of six histidines, i.e. His-tag, and glutathione S-transferase (GST). Normally, the commercial expression vectors contain the fusion partners. Some fusion partners may have an effect on the properties of the recombinant lectin. In this case, the fusion partners can be removed using a suitable protease such as thrombin. The addition of an adequate cleavage site can be done between the lectin and the fusion partner cDNAs during the cloning process if the available expression vector does not contain this cleavage site. If the fusion partners have no effect on the recombinant lectin properties, the fused lectin may be used without removing the fusion partners (Olausson *et al.*, 2011).

Several recombinant lectins from a variety of organisms (plants, mushrooms, animals, algae) have been produced by *E. coli* expression system (Oliveira, Texeira and Domingues, 2013). Impressive yields per litre of bacterial culture have been reported. For instance, the Griffithsin lectin (GRFT), a lectin from the red alga *Griffithia sp*., was expressed in *E. coli* and the total amount of the lectin in a 1 litre of bacterial culture was 819 mg. Of this total, 66%, that is, 542 mg was expressed in the soluble fraction. The expressed recombinant GRFT demonstrated similar biological activity in comparison to natural GRFT and presented exactly the same homodimeric structure (Giomarelli *et al.*, 2006).

Although *E. coli* expression systems have many advantages, it is important to point out some relevant limitations of these systems as well. *E. coli* is unable to produce glycosylated lectins, most of the lectins are expressed in inclusion bodies and these lectins are inactive so refolding is necessary. In addition, lectins with disulfide bonds are difficult to be expressed in a correct manner (Oliveira, Texeira and Domingues, 2013).

Bacteria are not capable of performing posttranslational modifications and this fact poses a major limitation when it comes to expressing recombinant lectins from eukaryotic origin. Several eukaryotic lectins undergo specific co- and posttranslational alterations in the native organisms, for example, (partial) *N*- and *O*-linked glycosylation or the formation of disulfide bonds. As discussed earlier, the glycosylation process is usually critical for the correct biological activity of a particular protein. Therefore, eukaryotic lectins expressed in *E. coli* systems should be assessed in terms of their stability and biological activity (Oliveira, Texeira and Domingues, 2013). For instance, the recombinant plant lectin, ricin B from *Ricinus communis*, was expressed in *E. coli* and the glycosylation was assessed as less stable than glycosylation on the native lectin (Frankel *et al.*, 1994; Ferrini *et al.*, 1995). Also, a lectin from the starfish *Asterina pectinifera* was recombinantly expressed in *E. coli* and did not seemed to form the necessary disulfide bonds which are required for hemagglutination activity, as well as high capacity to bind sugars (Kakiuchi *et al.*, 2002).

In order to reduce the cytotoxicity effect of overexpressed heterologous proteins, *E. coli* forms inclusion bodies, posing another drawback for the production of recombinant lectins. Most lectins are expressed in insoluble inclusion bodies so a technology has been developed to improve solubility. For example, the lectin *Allium sativum* leaf agglutinin (ASAL) is toxic to *E. coli* cells. However, the bacteria have been used to express this lectin as a fusion protein containing a particular peptide which enhances solubility and diminishes lectin cytotoxicity to the bacteria (Upadhyay *et al.*, 2010).

Functional lectins from the insoluble fraction containing the inclusion bodies can be successfully recovered (Longstaff *et al.*, 1998; Stancombe *et al.*, 2003; Luo, Zhangsun and Tang, 2005). The insoluble proteins are normally inactive and require refolding. However, a

small fraction refolds correctly and reestablishes the sugar-binding biological function (Streicher and Sharon, 2003).

Although *E. coli* expression systems have limitations regarding recombinant lectin production, these limitations have been overcome to a considerable degree, making *E. coli* an excellent host for recombinant lectins which do not require glycosylation.

1.1.7 Applications of Recombinant Lectins

Recombinant lectins have many advantages in relation to production and purification processes. Additionally, recombinant lectins presenting novel and improved properties such as mutated lectins (Yabe *et al.*, 2007) or peptide/protein fused lectin such as immunotoxins (Kreitman, 2006) can be produced to meet specific goals. In cancer research field, recombinant lectins have been developed for tumor biomarker and anti-tumor applications (Yang *et al.*, 2005; Oliveira *et al.*, 2009). Also, lectins produced by recombinant DNA technology have been investigated for infectious disease control to address anti-microbial (Kim *et al.*, 2007; Ling, Yang and Bi, 2010), anti-viral (Giomarelli *et al.*, 2006; Fouquaert *et al.*, 2009), and anti-insect (Luo, Zhangsun and Tang, 2005; Upadhyay *et al.*, 2010). Moreover, a set of novel technologies have been developed using recombinant lectins for cell profiling (Yim, Ono and Irimura, 2001; Maenuma *et al.*, 2008), lectin microarrays (Hsu, Gildersleeve and Mahal, 2008; Propheter and Mahal, 2011) and purification tag for recombinant protein production (Tielker *et al.*, 2006).

The direct identification of sugars on cell surfaces using lectins is a very promising technology for the investigation of cell surface glycosylation (see Figure 1.2). For instance, many mutants of the MAH lectin, a lectin from the legume *Maackia amurensis*, have been expressed in *E. coli*. These mutants have been successfully used to identify erythrocytes from different animal species (Yim, Ono and Irimura, 2001). Also, these lectins have successfully glycoprofiled cell lineage and differentiation stages of carcinoma, myeloid, fibroblastic and cells from melanoma (Maenuma *et al.*, 2008).

The use of recombinant lectins in a microarray format is a significant advancement in the analysis of cell surface glycosylation. Recombinant lectins expressed in bacteria have been demonstrated to be very useful for lectin microarrays (Hsu, Gildersleeve and Mahal, 2008). The fact that these lectins are not glycosylated by the bacterial expression system makes these recombinant proteins quite suitable for microarrays. Glycosylated lectins in microarrays may lead to false positive responses; for example, the binding of mannose-binding lectins present in mammalian cell lysates to high mannose carbohydrates attached to the immobilized lectins (Gupta, Surolia and Sampathkumar, 2010). For this reason, lectins sourced from plants present a considerable drawback for lectin microarray applications since these plant-derived lectins are mostly glycosylated.

Drug delivery is also a potential application of recombinant lectins Plattner et al. (2008) have investigated the use of lectins for site specific anti-tumor therapy. A recombinant plant lectin was used as carrier system for oral drug delivery. It was observed that the recombinant lectin remained integral and undamaged after digestion even for half an hour in simulated gastric and simulated intestinal fluid (Tremblay *et al.*, 2011).

Lastly, another important application of recombinant lectins is their use for facilitating the purification process of recombinant proteins. The lectins may be used as affinity tags in fusion constructs for a one-step protein purification process (Tielker *et al.*, 2006).

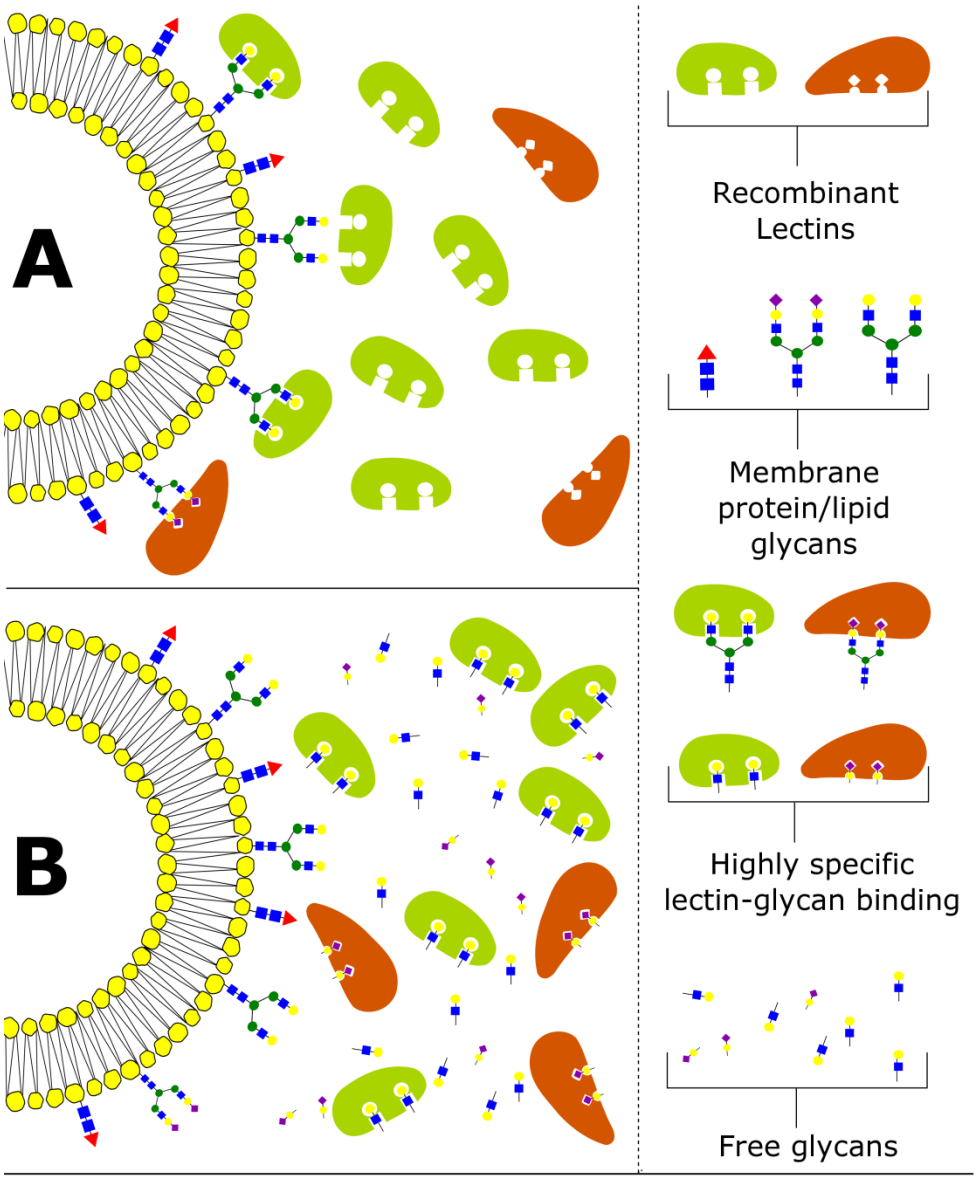

Figure 1.2: Schematic diagram illustrating the highly specific lectin-glycan interactions. A) Recombinant lectins specifically binding to glycan structures on cell membrane proteins and lipids, glycoprofiling cell surface. B) Free glycan structures inhibiting lectin interactions with carbohydrates attached on the cell membrane. The image was created with the aid of Inkscape 0.91.

1.2 The CHO cell

In 1919, Chinese hamsters were introduced first as laboratory specimens to replace mice for typing pneumococci. Later efforts to domesticate Chinese hamsters in the mid-20th century resulted in the development of spontaneous hereditary diseases owing to inbreeding. This fact encouraged researchers to investigate the genetics of hamsters (Yerganian, 1972, 1985) and it was discovered that the chromosome number of Chinese hamsters (2n = 22) was low, making these hamsters very useful for the study of radiation cytogenetics and tissue culture (Jayapal *et al.*, 2007).

In the late 1950s, in a study concerning the investigation of somatic cell genetics (Puck, Cieciura and Robinson, 1958) an ovary from a female Chinese hamster was isolated and cultured in cell culture plates. Researchers soon observed these cells were very resilient and they had relatively rapid generation times which made them very suitable for *in vitro* cultivation (Jayapal *et al.*, 2007).

1.2.1 CHO cells in the Biopharma Industry Context

The CHO cell line is the workhorse of the production of mammalian proteins, particularly at industrial scale. The human tissue plasminogen activator, the tPA, was the first recombinant protein to be commercially produced from mammalian cells (Deschenes, Finkle and Winocour, 1997). Since then, the annual global revenue of products sourced from CHO cell lines has increased to more than US$100 billion and the revenue continues to grow (Jadhav *et al.*, 2013, Jayapal et al. 2007). The main reason which has allowed CHO cells to be so successful is the incomparable adaptability which permits the growth of these cells at high densities when they are cultured in suspension which can be scalable to 10,000-L bioreactors and the use of serum free cultivation conditions (Jayapal *et al.*, 2007; Bandaranayake and Almo, 2014).

Chemically defined serum free media have been extensively improved in quality and availability. Such media are usually more cost effective since they do not contain or require

the supplementation with fetal calf serum . This fact makes these media safer, as the risk of viral and prion contamination from bovine serum is greatly reduced. Additionally, downstream processes can be simplified to a great degree, as chemically defined serum free media contain fewer protein contaminants (Bandaranayake and Almo, 2014). Furthermore, a scientific study which took place in 1989 investigated 44 human pathogens in CHO cells and it was concluded that the majority of these pathogens (such as human immunodeficiency virus (HIV), influenza, polio, herpes and measles) do not replicate in these cells. Therefore, CHO cells are ideal from a regulatory perspective (Jayapal *et al.*, 2007).

On the other hand, CHO cell line adaptability has some disadvantages. A production target must necessarily select a clone which has the required phenotypic features such as product quality and uniformity, doubling time and long-term viability under the conditions of a bioprocess. Phenotypic drifts, that is, alterations in the selected features of a clone, may occur even though the suitable CHO production clone has been identified (Jadhav *et al.*, 2013). However, the genomic variability of CHO cells has allowed the isolation of clones deficient in DHFR enzyme. This has resulted in a very effective way of selecting stable clones as well as the amplification of genes, thus increasing specific levels of productivity to a great degree (Jayapal *et al.*, 2007).

The highly adaptability feature of CHO cells associated with the knowledge and expertise gained over the decades and the extensive scientific research efforts to improve CHO production platforms, will surely keep CHO cells as the industry's premier workhorse for the production of therapeutic proteins at least in the near future.

1.2.2 Current monitoring tools for bioprocessing cell health

Most therapeutic proteins require critical and complex posttranslational modifications such as glycosylation, phosphorylation, and the formation of disulfide bonds (Zhao *et al.*, 2015). The regulatory agency looks at the profile of these modifications for the approval of a certain therapeutic protein production process and the agency requires that those PTMs are within a

range to ensure the quality of the expressed protein. As many quality attributes of a protein can be dictated by the process involving the cell cultivation, the monitoring and controlling of upstream bioprocesses are of paramount importance (Zhao *et al.*, 2015).

There has been substantial progress in the monitoring of bioprocesses and most of the methods measure physical and chemical factors such as cell concentration and nutrient levels in order to track unexpected changes and control the process if it is needed (Zhao *et al.*, 2015). However, methods which investigate cell surface glycosylation and the correlation with therapeutic protein glycosylation have not been developed, despite the fact that changes in glycosylation patterns have been shown to be linked to the onset of abnormal or unhealthy state of the cell (Veiseh et al. 2014; Krasnewich 2014; Gorelik et al. 2001; Patsos et al. 2009; An et al. 2009).

Currently, spectroscopic based methods have been used as tools to monitor bioprocesses. Methods such as Near infrared (NIR) spectroscopy and Mid infrared (MIR) spectroscopy are examples of *in situ* analytical techniques which have been used to monitor the culture composition of CHO cells and NIR has been implemented in industrial settings such as Eli Lilly, Novo Nordisk and AstraZeneca (Forcinio, 2003). 2D fluorometry, Electronic nose and Dielectric capacitance are also *in situ* spectroscopic analytical techniques to monitor mammalian cell processes, including CHO cells (Teixeira *et al.*, 2009). These methods require interpretation of the spectral data. This means that specific models (such as chemometric models) are needed to extract meaningful information, as the multidimensional nature of the data cannot be associated straightforwardly to a given target bioprocess variable (Teixeira *et al.*, 2009). Although some of these methods provide information on cell viability and the protein of interest, the glycosylation state of cell surface as a parameter to monitor glycan patterns to identify early signs of cell stress is not obtained. Therefore, these methods fail to investigate the correlation of cell surface glycosylation and the glycosylation profile of the therapeutic protein, which is a critical quality attribute (Zhao *et al.*, 2015).

Flow cytometry is a powerful tool which has been used in laboratories to investigate cells for several years. Thus, this technique can potentially be used to analyse cell physiology for the understanding and prediction of the process kinetics for tighter control and improvement of

the bioprocessing of therapeutic proteins in industrial settings (Kuystermans, Mohd and Al-rubeai, 2012). The multidimensional information obtained from flow cytometry on cell population can include cell size, viability. By using specific staining, information on cell surface glycosylation, through the use of fluorescent lectins; (see Section 1.1.7), intracellular proteins, DNA cell cycle and apoptosis can be obtained as well (Zhao *et al.*, 2015; Kuystermans, Avesh and Al-rubeai, 2016).

Although flow cytometry analysis is commonly used as an off-line tool, the potential use of this technique as an on-line monitoring system has been demonstrated (Zhao, Natarajan and Srienc, 1999; Sitton and Srienc, 2008; Broger *et al.*, 2011; Kuystermans, Avesh and Al-rubeai, 2016). As the automation of flow cytometry for monitoring of biopharmaceutical manufacturing processes has become more promising, the on-line analysis of glycosylation patterns on cell surface presents itself as a relevant alternative to be used as one of the main parameters to identify glycosylation changes which may indicate early signs of cell stress leading to compromises in the quality of the protein of interest or even cell death. Therefore, such early signs may then be mathematically associated with culturing conditions to develop a predictive system to control the bioprocess more tightly.

It is the purpose of this present research work to investigate the relationship between the culturing parameters and cell surface glycosylation alterations to address fundamental questions which can potentially establish the foundations for the development of a bioanalytical tool based on cell surface glycosylation analysis. This tool might not only be able to identify early signs of cell stress but also provide data to build a controlling system to act upon, accordingly ensuring the bioprocess trajectory is within expectation.

1.2.3 Glycosylation in CHO Cells

CHO cells are able to synthesize a number of complex and oligomannosyl *N*-glycans with few hybrid structures (Lee *et al.*, 2001), mucin *O*-glycans containing up to four monosaccharides (Sasaki *et al.*, 1987), and *O*-fucose (Moloney *et al.*, 2000), *O*-glucose (Moloney *et al.*, 2000), *O*-

mannose (Patnaik and Stanley, 2005) glycans, and polysialic acid which has been found as a minor portion of glycoproteins (Muhlenhoff *et al.*, 1996; Hong *et al.*, 2004).

GM_3 is the major glycolipid synthesized in CHO cells (Stanley, Sudo and Carver, 1980; Warnock *et al.*, 1993). Heparan sulfate and chondroitin-sulfate proteoglycans are also found in CHO cells (Esko, Stewart and Taylor, 1985). CHO cells lack the expression of glycosyltransferases which transfer α1,2-, α1,3-, or α1,4-linked fucose (Howard *et al.*, 1987), β1,6-linked *N*-acetylglucosamine (GlcNAc) to generate core 2 *O*-glycans (Sasaki *et al.*, 1987; Bierhuizen and Fukuda, 1992), sialic acid α2,6-linked to Gal (Sasaki *et al.*, 1987), or the bisecting GlcNAc (Campbell and Stanley, 1984). Sulfotransferase activities associated with the formation of sulfated glycolipids or sulfate *N*- or *O*-glycans are not present in CHO cells either (Brockhausen, Vavasseur and Yang, 2001).

1.2.4 Glycosylation alterations during cell stress

The majority of proteins synthesized in eukaryotic cells are altered during or just after translation. As mentioned in Section 1.1, these alterations are named post-translational modifications (PTMs) and they are of covalent nature which have the purpose of providing an extra level of regulation for proteins and to provide proteins with a selective ability to be involved in different processes (Walsh, Garneau-tsodikova and Gatto, 2005; Freeze and Schachter, 2009; Boscher, Dennis and Nabi, 2011). Glycosylation is one of the most critical PMT and proteins are glycosylated in the endoplasmic reticulum (ER) and Golgi apparatus. The ER also functions as a protein control quality unit by sorting proteins which have not been properly folded (Zhang and Kaufman, 2006).

The secretory pathway initiates in the ER and terminates at the trans-Golgi. This pathway at normal conditions provides properly folded and glycosylated proteins to the surface of the cell. Such activity is of paramount importance for the development and homeostasis of eukaryotic cells and cell-to-cell communication in multi-cellular organisms (Dennis, Lau and Nabi, 2009a; Boscher, Dennis and Nabi, 2011). ER quality control function ensures that

misfolded or unfolded proteins are retained and recycled so that only properly folded proteins are sent to Golgi apparatus (Zhang and Kaufman, 2006). Disruptive alterations in calcium homeostasis or redox status, glucose deprivation, overexpression levels of proteins, altered glycosylation and the expression of misfolded proteins are some examples of stimuli which can impose stress to the ER. Therefore, situations which involve ER stress can profoundly impact protein glycosylation (Ruddock and Molinari, 2006) by modifying regulation of pathways related to the unfolded protein response (UPR), endoplasmic reticulum associated protein degradation (ERAD) and secretion of proteins for instance (Zhang and Kaufman, 2006; Chakrabarti, Chen and Varner, 2011).

In the context of industrial bioprocessing involving CHO cells, sodium butyrate (NaBu) has been extensively used to increase the expression levels of recombinant proteins. However, NaBu has effects on the quality of glycoprotein such as elevated heterogeneity, decrease in vivo biological activity and alterations of the glycosylation of the expressed protein (Sung *et al.*, 2004). Furthermore, sodium butyrate can also inhibit cellular growth and induce apoptosis (Kim and Lee, 2002). Changes in culture conditions such as temperature and cultivation mode can affect glycosylation patterns of glycoproteins. For instance, *N*-linked glycans on secreted human placental alkaline phosphatase, a glycoprotein which was produced on CHO cells, showed alterations when the temperature was reduced. Also, glycosylation was altered when CHO cells were cultivated in microcarrier culture (Nam *et al.*, 2008).

Many scientific studies have reported glycosylation alterations on the expressed proteins (Werner, Kopp and Schlueter, 2007; Zheng, Bantog and Bayer, 2011; Shi and Goudar, 2014; Zheng *et al.*, 2014; Wada, Matsui and Kawasaki, 2019). However, there is a relatively low number of reported studies on cell surface glycosylation changes. Grainger & James (2013) set up a series of experiments to investigate the correlation between cell surface glycosylation and expressed protein glycosylation. It was observed that monoclonal antibody galactosylation and CHO cell surface galactosylation were significantly correlated in a substrate-controlled variation experiment. Additionally, the researchers demonstrated that it is possible to predict and control *N*-glycan glycosylation process of a secreted recombinant glycoprotein based on measurements obtained from cell surface glycans using lectins. The

findings of this investigation indicate that cell surface glycans may be used to monitor the health of the bioprocessing cell as well as the glycosylation of the expressed glycoprotein.

1.3 Flow cytometry: basic principles

Flow cytometry is a powerful technique both in research and clinical settings for the definition of cellular characteristics or particles. Light scattering and fluorescence emission are the physical phenomena exploited by the technique as cells or particles in suspension are interrogated by an optical system. An electronics system transforms the optical data into a digital dataset which can be visualized and interpreted. Since its first development about 60 years ago, flow cytometry measurement capability has been expanded from measuring the relative cell size parameter only (Coulter Counter) to 18 parameters measured simultaneously (for example, Becton Dickinson's FACSAria III) (Wilkerson, 2012; Adan *et al.*, 2017). The most common applications of flow cytometry are the detection of membrane, cytoplasmic and nuclear antigens, whole cells and cellular components, and the analysis of the DNA cell cycle and cell proliferation (Adan *et al.*, 2017).

1.3.1 Fluidics system

The fluidics system has the purpose of transporting particles suspended in a fluid stream to the laser beam for interrogation. The interrogation process is optimised when the stream transporting the particles is placed in the center of the laser beam and when only one particle is moved through the laser at a given moment (Graves and Pearlson, 2013).

The accomplishment of the optimal interrogation process is achieved by the injection of the sample into a stream of sheath fluid in the flow chamber. The flow chamber design focuses the sample core in the center of the sheath fluid where the laser beam interacts with the particles. Laminar flow is the principle governing the separation of the sample core from the

sheath fluid whose flow rate accelerates the particles and restricts them to the center of the sample core. Such process is known as hydrodynamic focusing and it is illustrated in Figure 1.3.

The sample pressure is always greater than the sheath fluid pressure. The sample flow rate is then regulated by controlling the sample pressure in relation to the sheath fluid pressure. By increasing the sample pressure the flow rate increases widening the sample core. This causes more cells to enter the stream in a given moment. However, by increasing the number of cells entering the stream, it could increase the number of cells passing the laser beam off-center; thereby, the cell is sub optimally interrogated. Nevertheless, this may be appropriate for some applications. For instance, qualitative measurements for immunophenotyping can be taken at a higher flow rate. However, because the cells are less in line in the wider core stream, the data obtained is less resolved, but it is quicker to acquire. On the other hand, a lower flow rate reduces the width of the sample core, restricting the cells to a smaller area. Therefore, the vast majority of cells is interrogated in the center of the laser beam. This ensures the light shining on the cells and emitted from them is more uniform. DNA analysis requires high resolution; thereby, a lower flow rate is generally used in this application (Shapiro and Telford, 2009; Wilkerson, 2012; Adan *et al.*, 2017).

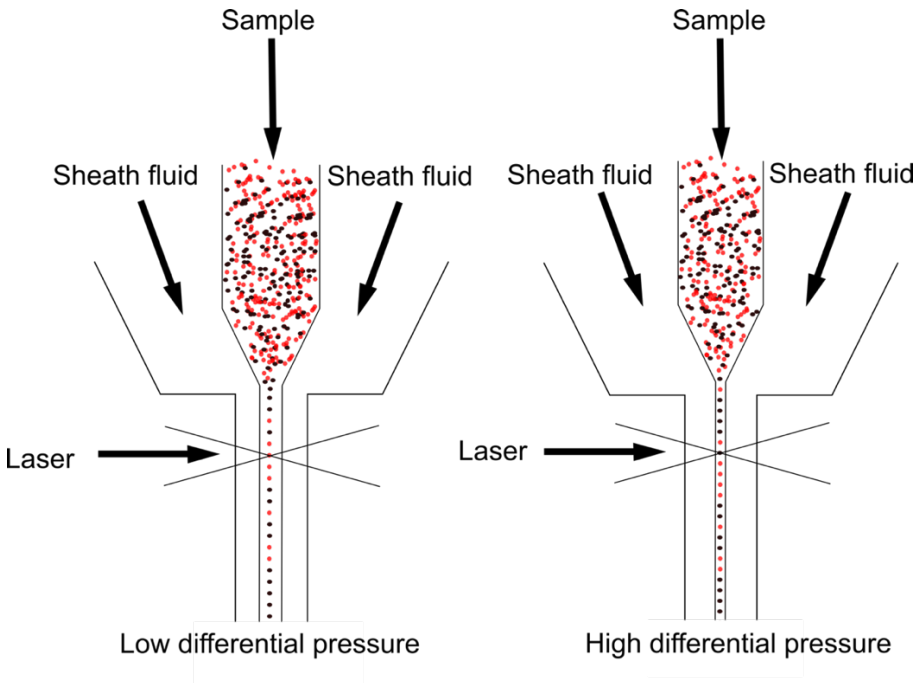

Figure 1.3: Hydrodynamic focusing of the sample core.

1.3.2 Generation of scatter light and fluorescence

By hydrodynamic focusing, cells or particles are transported to the interrogation point at which a laser light shines. In order to understand what happens to laser light and how signals are processed as the light strikes a cell/particle, the concept of light scattering and fluorescence is discussed first.

Light scattering is characterised by the deflection of incident laser light when it encounters a particle. This phenomenon depends on the physical properties of a particle such as its size and internal complexity. The cell membrane, nucleus, organelles (or any granular material in the cell), the cell shape and the topography of its surface are factors which can affect light scattering (Wilkerson, 2012).

Therefore, forward-scattered light (FSC) provides information on the relative cell-surface area or size. FSC measures most of the diffracted light being detected by a photodiode just off the axis of the incident laser beam in the forward direction. (Figure 1.4). FSC is a suitable method

of detecting particles based on a given size without collecting any information of their fluorescence. Whereas, side-scattered light (SSC) provides information on the granularity or internal and external complexity of the cell. SSC measurements are mostly taken from refracted and reflected light that occurs at any cell interface where a change in refractive index takes place (Figure 1.4). The collection of SSC is at approximately 90 degrees to the laser beam by a set of lens and then the light is redirected by a beam splitter to the appropriate detector (Wilkerson, 2012; Adan *et al.*, 2017).

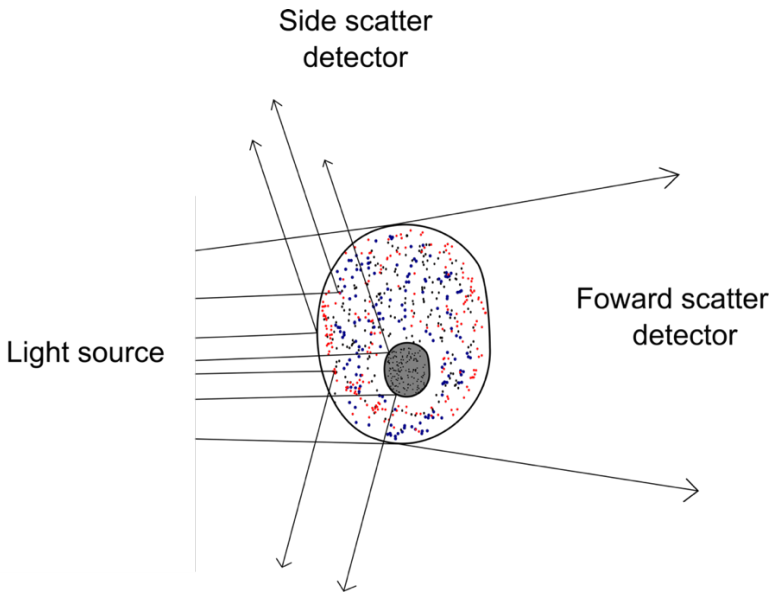

Figure 1.4: Illustration of light scattering properties of a cell.

Fluorescence is characterised by the emission of light from a fluorescent compound due to the absorbance of light energy. A compound absorbs light over a range of wavelengths which depends on the chemical composition of the compound. The absorption of light excites an electron in the compound, leading this charged particle to a higher energy level. However, the excited electron quickly returns to its original energy state, thereby releasing the excess energy in the form of a photon of light (Wilkerson, 2012).

Absorption spectrum is the range of wavelength which can excite a particular compound, whereas emission spectrum is the range of emitted wavelengths of this compound. Since more energy is consumed during the light absorption than its emission, the wavelengths of emitted light are longer than the absorbed ones.

As a result, more than one fluorescent reagent can be used simultaneously and excited at the same wavelength as long as the peak emission wavelengths are not very close to each other. For instance, fluorescein isothiocyanate (FITC) and phycoerythrin (PE) can be excited at 488 nm and both emission peaks can be easily identified (Figure 1.5). The intensity of fluorescent signal detected is proportional to the number of fluorescent molecules on the cell/particle.

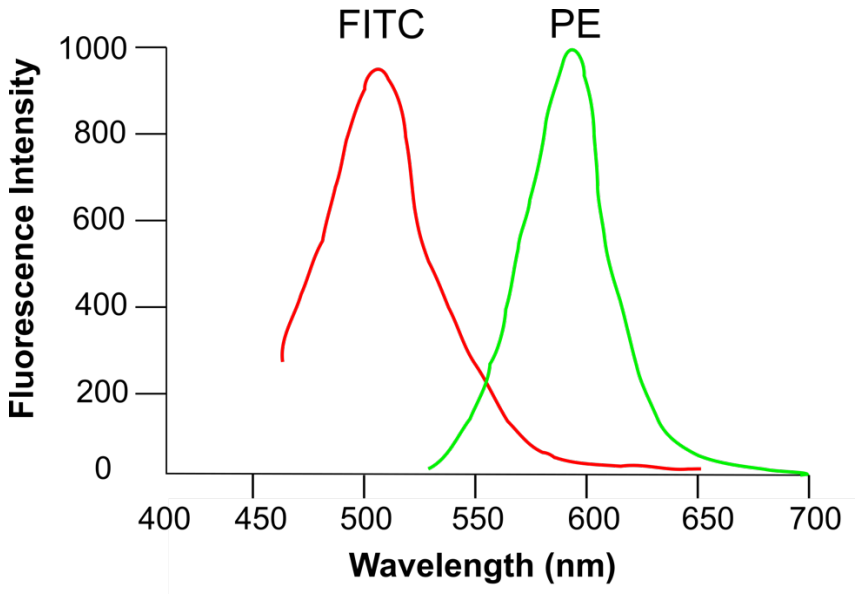

Figure 1.5: Emission spectrum of FITC and PE excited at 488 nm.

A monoclonal antibody conjugated with a fluorescent reagent is very useful in the identification of a particular cell type. Thus, specific antigenic markers of the cell are used to achieve the cell type identification (Figure 1.6). As a consequence, heterogenous cell population can be distinguished into separate subpopulations by employing different fluorochromes. The combination of the data extracted from FSC and SSC channels with the

staining pattern data of each subpopulation can be used to provide information on which cells are present in a sample and to quantify their relative percentages. In addition, modern flow cytometers can sort the cells if required.

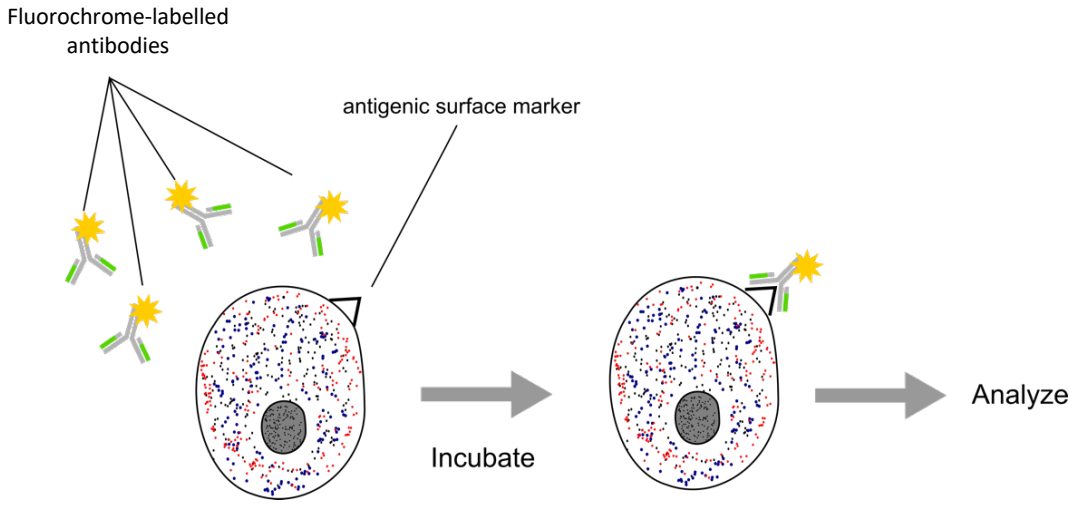

Figure 1.6: Interaction of specific fluorochrome-labeled antibodies with cell surface antigens markers.

1.3.3 Optics system

The optics system is characterised by the excitation and collection optics. Laser and lenses that are employed to shape and focus the laser beam compose the excitation optics, whereas the collection optics is composed of a set of lenses which collect light emitted from the particle due to the interaction with the laser beam. Also, a system of optical mirrors and filters composes the collection optics to direct specified wavelengths of the collected light to designated optical detector channels. These functions are achieved by the design of the optical bench which provides a fixed position of the light source and the excitation and collection optics. Therefore, the laser intercepts the sample stream in a consistent manner (Adan *et al.*, 2017).

The emitted SSC light and fluorescence signals resulting from the interrogation of a cell or particle at the laser beam are diverted to the photomultiplier tubes (PMTs) and the FSC signals are collected by a photodiode. All of the signals are directed to their designated detectors by a collection of mirrors and optical filters. Fluorescence signals, which are generally weak, are detected by PMTs. An optical filter placed in front of the PMT enables a detector channel to specifically detect the fluorescence emitted by a particular fluorescent reagent, since the filter allows only a narrow range of wavelengths to reach the channel (Figure 1.7) (Adan *et al.*, 2017).

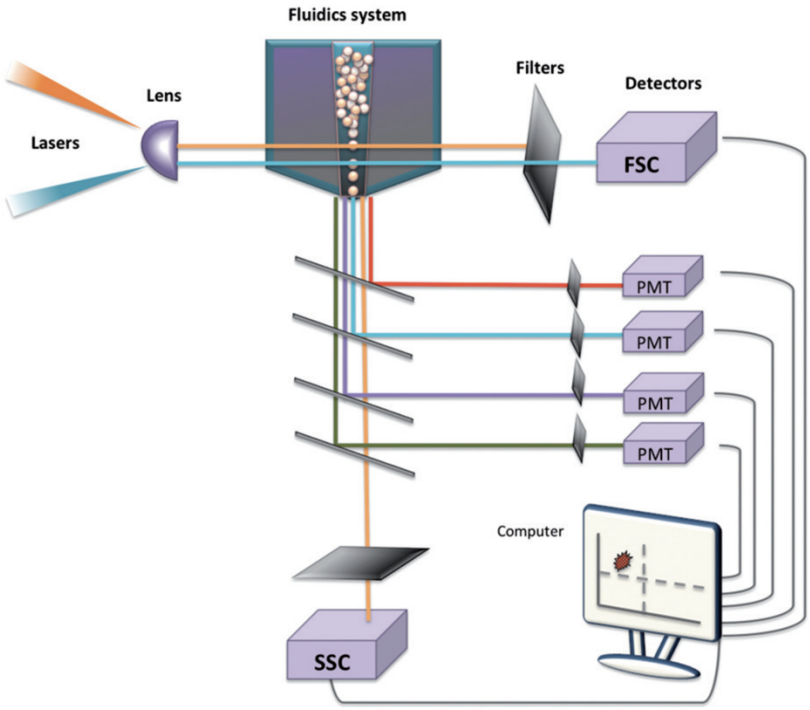

Figure 1.7: Schematic diagram of an optical bench of a typical flow cytometer. Figure was extracted from Adan et al. 2017.

1.3.4 Electronics system: signal detection and processing

As particles reach the interrogation point, thus passing through the laser beam, signals of light are generated and then converted into electronic signals (voltages) by photodetectors. The voltages are then assigned a channel number on a data plot. Generally, photodetectors are of two types: photodiodes and photomultiplier tubes (PMTs). Stronger light signals such as FSC signal are detected by the photodiode since it is less sensitive than the PMTs. On the other hand, weaker light signals such SSC and fluorescence are detected by the PMTs (Wilkerson, 2012).

An electronic signal in the form of a voltage pulse is created whenever a particle passes through the laser beam and light scattering and fluorescence occur. Once the PMT or the photodiode is hit by light signals or photons, the photodetectors convert them into a proportional number of electrons which are multiplied. This, increases the electrical current which travels to the amplifier, where it is converted into a voltage pulse. The maximum amount of light scattering and fluorescence is achieved when the particle is at laser beam center; therefore, generating the highest point of the pulse. However, the pulse drops to its baseline level as the particle leaves the laser beam (Figure 1.8). The digitalisation of the voltage pulse is achieved by an Analog-to-Digital Converter (ADC). The height of a voltage pulse is the maximum amount of current generated at the PMT, the width the pulse gives the interval it occurs, and the area is the integral of the pulse. Thus, signal intensity can be measured by either the height or the area of the pulse (Adan *et al.*, 2017).

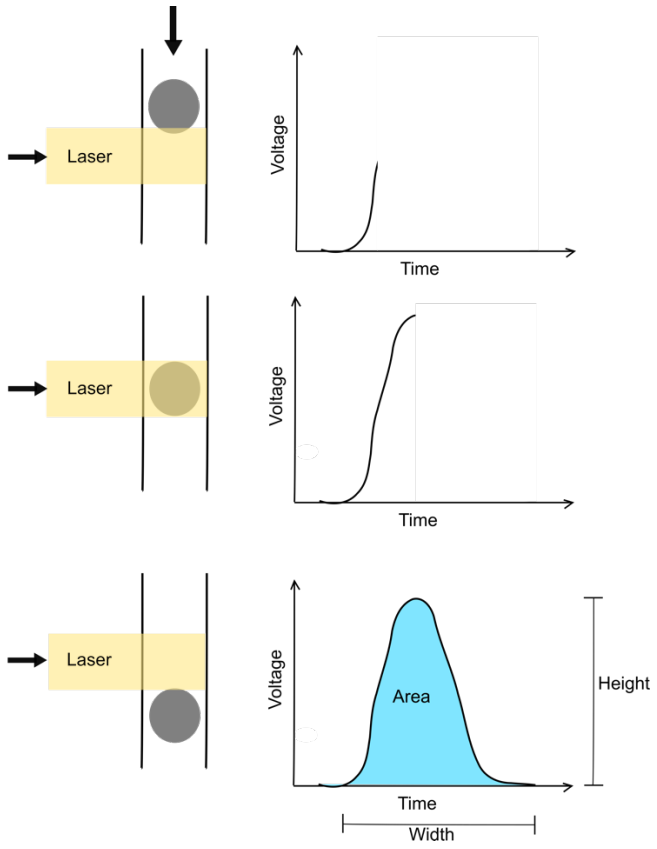

Figure 1.8: Generation of a voltage pulse as a particle passes through the laser beam.

The PMT voltage, the amplifier gain and the number of photons which are detected determine the voltage pulse size. Therefore, signals can be amplified by applying a voltage to the PMTs causing an increase in the electrical current, or by elevating the amplification gain. The logarithmic amplification is generally used for discrimination of negative from dim positive signals, while the linear amplification is generally used for the amplification of scatter and fluorescence parameters (Adan *et al.*, 2017).

1.4 Flow cytometric DNA cell cycle analysis

The cycle of the eukaryotic cell can be divided into four distinct stages or phases: G1, S, G2 and M (Figure 1.9). In G1, growth cell takes place before the S phase starts by initiating DNA

synthesis. In G2, cells grow again prior to cell division which takes place in M or mitotic phase. The daughter cells produced by this division may only successfully survive provided (i) that each phase of the cell cycle takes place in the correct sequence (ii) is completed prior to the initiation of the next phase and (iii) that each phase is faithfully processed (Tate and Ko Ferrigno, 2006).

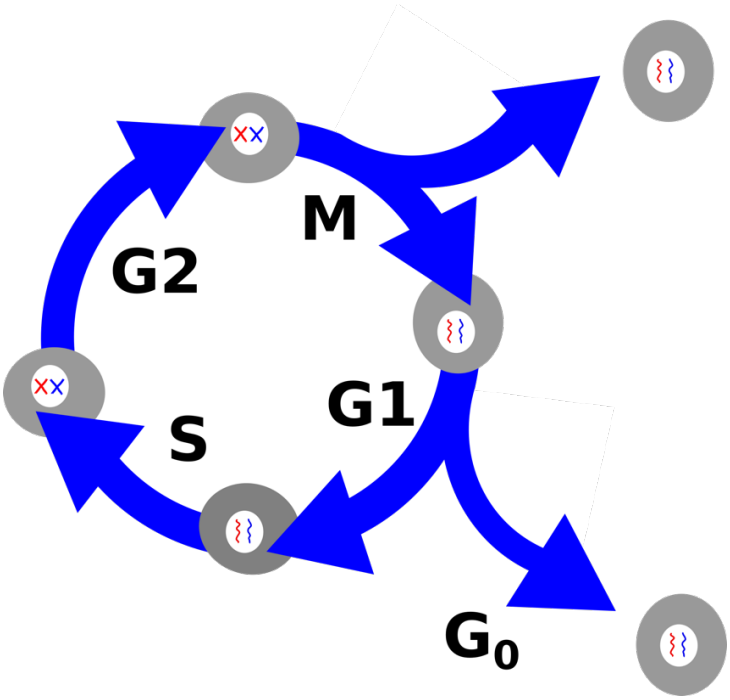

Figure 1.9: Cells in G1 phase might decide to exit the cell cycle and initiate a quiescent Go phase. Cells remaining in G1 phase undergo the duplication of DNA (S phase) before entering mitosis.

Flow cytometry allows the quantitative measurement of the nuclear DNA content through the use of fluorescent reagent which binds stoichiometrically to the DNA. In other words, the stained cellular material incorporates an amount of fluorescent dye proportional to the amount of DNA. Thus, the height of the electronic pulse generated is proportional to the total fluorescence emission from the cell, allowing the identification of subgroups of cells based on their DNA content (cell cycle phases) (Nunez, 2001).

In a proliferating cell population, three distinct DNA cell subpopulations can be identified through flow cytometry: Go/G1, S, G2/M. The data is presented as cellular DNA content frequency histograms (Figure 1.10) (Nunez, 2001).

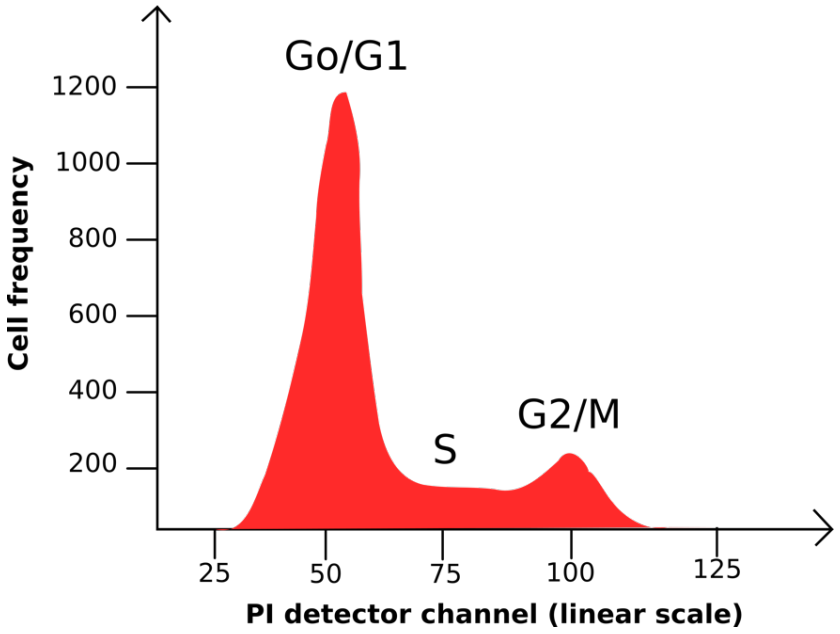

Figure 1.10: Illustration of a DNA histogram obtained from staining cells with propidium iodide (PI), a common fluorescent reagent for DNA staining.

Cells in Go/G1 all have a uniform DNA content, as do cells in G2/M. Since the latter cells have twice as much DNA than Go/G1 cells, the G2/M peak is located at twice as much the fluorescent value of Go/G1 peak and the S cells are in between the two peaks. Therefore, by identifying the first peak (Go/G1), the remaining DNA subpopulations can be recognized. The coefficient of variation (CV) of the mean value of the fluorescence related to DNA content of Go/G1 subpopulation is a reflection of the accuracy of DNA content measurement. Thus, a CV lower than or equal to 6% ensures a great level of accuracy of the measurement of DNA content (Pozarowski and Darzynkiewicz, 2004).

1.5 Current technologies on cell surface glycoprofiling and challenges

Several lectin-based approaches have been developed for cell surface investigation although monoclonal antibodies have also proved to be invaluable tools for the analysis of complex glyconjugates. In general, monoclonal antibodies bind terminal components, thus limiting the utility of these proteins in the analysis of such terminal components. Conversely, lectins may bind to both core and terminal glycan structures. Currently, Mass Spectrometry, immunohistochemistry and flow cytrometry are techniques widely used for the characterization of cell surface glycosylation using lectins (Chen *et al.*, 2007).

Mass Spectrometry (MS) is a powerful technique which is applicable to the analysis of complex glycans was first accomplished with the development of a tandem MS technique (Hirabayashi, 2008). A characteristic degradation pattern can be associated with individual glycans allowing their effective differentiation. MS provides high accuracy (resolution) for both confirmation and estimation of glycan structures. However, this technique requires previous treatments which involves the liberation of glycans from proteins and lipids prior to the modification with an appropriate labelling reagent, such as 2-aminopyridine (Hirabayashi, 2008). Generally, these pre-treatments are time-consuming and the resulting N-glycan pool can contain intra- and extracellular proteins (Hamouda *et al.*, 2014).

Immunohistochemistry technique became very popular during the 1980s to investigate the distribution of several markers in normal and diseased tissues. This technique uses antibodies as reagents for the detection of the cell or tissue localization of a specific antigen, which is identified by a label. By using microscopy, this label can be identified. Lectin histochemistry was then developed using the basic concept underlying immunohistochemistry; thus, labelled lectins were used to detect their binding to carbohydrate structures (Brooks, 2017). A major application of lectin histochemistry has been the investigation of alterations in cellular glycosylation as normal cells become malignant, and alterations associated with cancer progression (Brooks, 2017). Although lectin histochemistry is very powerful, the paraffin-embedding methods which are normally used for tissue fixation can make carbohydrates in glycoproteins inaccessible owing to protein denaturation. In addition, glycolipids can be lost

during the fixation process. Thus, methods which allow the use of unfixed biological material can be more advantageous. Furthermore, although great advancements have been made towards developing high throughput lectin histochemistry analyses (Pilobello, Slawek and Mahal, 2007; Tateno *et al.*, 2007), the overall process can still be lengthy and expensive due to multiple steps and costly reagents (Chen *et al.*, 2007).

As mentioned earlier, flow cytometry is a very powerful technique which allows the rapid extraction of light scattering and fluorescent based information from cell by cell (see section 1.3). The technique allows the use of fixed or unfixed cells which can be incubated with labelled lectins for the detection of lectin binding to glycans on the cell surface. Therefore, this technique allows the analysis of living cells in a rapid manner unlike mass spectrometry and lectin histochemistry. In addition, cells can also be stained with other fluorescent reagents which provide information on different aspects of the cell population under investigation such as live/dead dyes. Thus, this information can be combined with the lectin binding pattern (Batisse *et al.*, 2004; Stanley and Sundaram, 2014).

As the use of multiple dyes greatly increases the level of complexity of flow cytometry data, several automation tools to analyse the data have been developed, providing consistency to the data analysis task (Rahim *et al.*, 2018; Conrad *et al.*, 2019; Montante and Brinkman, 2019). Although there are many scientific reports on the glycosylation of the cell surface using flow cytometry, those reports have not explored the possibility of automating flow cytometry data analysis using computer languages (Batisse *et al.*, 2004; Stanley and Sundaram, 2014). Furthermore, possibly due to the increased complexity of the data analysis, scientific investigations which employ automation of flow cytometric data analysis of DNA cell cycle combined with cell surface glycosylation have not been reported yet. Studies have shown cell surface glycosylation changes associated with the cell cycle (Slawson *et al.*, 2005; Chen *et al.*, 2010; Ozlu *et al.*, 2015), thus the discrimination of the DNA populations in a flow cytometry analysis provides a higher quality of data while increasing the knowledge of the relationship between DNA cell cycle and glycosylation changes in the cell surface. In summary, the development of a rapid automated cell surface glycosylation combined with DNA analysis using flow cytometry would greatly contribute to the field of glycobiology.

1.6 Research aims

This research work mainly aims to develop a rapid and automated bioanalytical methodology to monitor the bioprocessing cell health by using fluorescent recombinant lectins for probing the cell surface and using DNA fluorescent reagents to identify cell surface glycosylation changes across the cell cycle.

As discussed earlier, there have not been scientific reports combining DNA cell cycle with lectin-based cell surface analysis using flow cytometry and data analysis automation to consistently monitor changes in cell surface glycosylation. In addition, the biopharmaceutical industry lacks a technology which can provide information on cell health during the bioprocess within a shorter period of time. Therefore, this high throughput methodology is going to allow the monitoring of the bioreactor process step in a more consistent and efficient way, dramatically reducing operational costs associated with batch loss due to diseased cells.

Chapter 3 describes the methodology developed to prepare the cells for analysis through flow cytometry and how to filter the data in order to look at only viable single cells. Chapter 4, particularly section 4.9, presents multiple data visualization formats combined with the statistical analysis facilitating the understanding of the complex flow cytometric data obtained.

2 Materials

2.1 Strains of *Escherichia coli*

Table 2.1: *E. coli* strains and details

Strain	Use in project	Features	Source
JM109	Used for making competent cells and for initial small scale protein expression.	Enhanced for high quality miniprep DNA. The *rec*A1 mutation improves insert stability. Appropriate for routine cloning.	Stratagene[a]
KRX	Used for higher level of protein expression.	Engineered for optimised controlled protein expression. The *rec*A- mutation minimizes undesirable recombination events. The *omp*P and *omp*T mutations reduce the proteolysis of overexpressed proteins.	Promega[b]

a - JM109; b - KRX 2013

2.2 CHO-K1

Table 2.2: CHO-K1 details

Cell line description:	A subclone of the parental CHO cell line, which was derived from the ovary of an adult Chinese hamster.
Species:	*Cricetulus griseus*, hamster, Chinese
Tissue of origin:	Ovary
Celltype:	Epithelial
Growth mode:	Suspension
Biosafety level:	1
Source:	Dr. Niall Barron, NICB

2.3 Microbiological Media

The chemical solutions used for the preparation of the microbiological media were all ACS grade and were supplied by Sigma-Aldrich unless otherwise stated. An autoclave was set up at 121 °C for a 20 minute cycle to sterilize the media. Distilled water (dH_2O) was from a Mili0Q® Academic system with a MILLIPAK™ 0.22 µm filter.

Luria Bertani Broth (LB)

Tryptone	10 g/L
NaCl	10 g/L
Yeast Extract	10 g/L
pH	7.0

The pH was adjusted to 7.0 prior to sterilisation using a NaOH solution and brought to the correct volume using dH_2O. In order to produce solid agar plates, 15 g/L of agar was added prior to sterilization.

Terrific Broth (TB)	
Tryptone	12 g/L
Yeast Extract	24 g/L
Glycerol	4-8 mL/L

Distilled H_2O was added to bring the volume to 900 mL before autoclaving. After allowing it to cool, 100 mL of 1M potassium phosphate buffer (see section 2.5), was added aseptically and the pH was adjusted to pH 7.4.

2.4 Cell Culture Medium

Table 2.3: CHO K1 Cell Culture Supplementation

Ingredients	Quantity (mL)	Functions
BalanCD CHO Growth A Medium (Irvine Scientific 91128-1L)	960	Base medium
Penicillin – Streptomycin (Sigma P4333)	10	Antibiotics
12.6% PVA (Sigma P8136) in PBS (w/v) (Sigma D8537)	20	Anti-foaming and anti-aggregating agent
L – Glutamine 2mM (Sigma G7513)	10	Energy source

Cell culture media were supplemented aseptically (with the aid of a biological cabinet) with all the ingredients except for L – Glutamine (and this partially supplemented media was stored

at 4 °C. As L-Glutamine is quite unstable when it is in media, the required amount was added to the partially supplemented media on the day cells were subcultured.

2.5 Buffers and Solutions

Potassium Phosphate Buffer (1.8 M)

KH_2PO_4	23.1 g
K_2HPO_4	125.4 g
pH	7.4

The required volume of dH_2O was added to bring the volume to 1 L before autoclaving.

PBS (10X)

Na_2HPO_4	10.9 g/L
NaH_2PO_4	3.2 g/L
NaCl	90 g/L

In order to make TBS, Triton X-100 was added to PBS 1X to a concentration of 0.1 % (v/v).

TBS Buffer

Tris	20 mM
NaCl	150 mM
pH	7.6

An HCl solution was used to adjust the pH. $CaCl_2$ was added to a final concentration level of 1 mM. In order to make TBST, Tween 20 was added to a final concentration level of 0.1 % (v/v).

Then, BSA was added to TBST to a final concentration level of 5% (w/v) to make western blot blocking buffer.

SDS-PAGE Buffer (5X)

Tris	15 g/L
Glycine	72 g/L
SDS	5 g/L
pH	8.3

The buffer was diluted using dH_2O to a 1X running buffer solution.

SDS Sample Buffer (6X)

4X Tris-Cl, pH 6.8/SDS	7 mL
Glycerol	3 mL
SDS	1 g
DTT	0.93 g
Bromophenol Blue	1.2 mg

Distilled water was added to bring to 10 mL, if necessary. Aliquots of 0.5 mL were prepared and stored up to 6 months at -80 °C.

4X Tris-Cl/SDS Buffer

Tris base	6.05 g
dH_20	40 mL
pH	6.8

The pH was adjusted to 6.8 with a 1 N HCl solution. Then 0.4 g of SDS was added followed by dH_2O to 100 mL total volume.

Coomassie blue stain solution

dH_2O	50 % (v/v)
Methanol	40 % (v/v)
Acetic Acid	10 % (v/v)
Coomassie blue	0.25 % (w/v)

Coomassie blue destain solution

dH_2O	45 % (v/v)
Methanol	45 % (v/v)
Acetic Acid	10 % (v/v)

Lysis Buffer

NaH_2PO_4	50 mM
NaCl	0.5 M
Imidazole	10-250 mM
pH	8.0

Western Blot Semi Dry Transfer Buffer (SDTB)

Tris base	5.8 g
Glycine	2.9 g
Methanol	200 mL
SDS	0.37 g

SDS solution

SDS	10 % (w/v)

Distilled water was added to 1 L total volume and the solution stored at 4 °C.

Citrate Buffer

Sodium Citrate (0.1 M)	3.63 mL
Citric Acid (0.1 M)	1.37 mL
dH_2O	5 mL
pH	5.5

TMB Solution (10 mL)

Citric Acid	1.37 mL of 0.1 M Stock
Sodium Citrate	3.63 mL of 0.1 M Stock
dH_2O	5 mL

One TMB tablet (T3406) (1mg) was dissolved in 200 µL of dH_2O and added to 9.8 mL TMB solution. Two microliters of H_2O_2 was added immediately before use.

PBS/CaMg

A volume of 900 mL of distilled water was added to a beaker with a stir bar, then the following compounds were added: 8 g of NaCl, 0.2 g KCl, 1.44 g of Na_2HPO_4, 0.25 g of KH_2PO_4, and 0.2 g of hexahydrate $MgCl_2$. Once these compounds were fully dissolved with stirring, the volume was adjusted to 1 litre with distilled water. HCl (2N) was then added drop by drop to adjust the pH to 7.2. The solution was then filtered through a 0.22 µm filter and stored indefinitely at 4°C (Stanley and Sundaram, 2014).

7-Aminoactinomycin D

A volume of 50 µL of dimethyl sulfoxide (DMSO) was added to 1mg of 7-Aminoactinomycin D (7-AAD; ThermoFischer Scientific A1310) to promote the dissolution. Then 950 µL PBS/CaMg was added to obtain 1 mg/mL solution. This solution was then kept protected from light and stored for up to 6 months at 4°C (Stanley and Sundaram, 2014).

3 Methods

3.1 Antibiotics and IPTG

Stock solutions of ampicillin were prepared in dH_2O at a concentration level of 100 mg/mL and stored at -20 °C. LB agar plates and broth cultures were prepared at 100 µg/mL of ampicillin (working concentration). Likewise, stock solutions of IPTG at 100 mg/mL were made and used at a working concentration level of 100 µg/mL.

3.2 Storing and culturing of bacteria

Bacterial stocks were stored at -80 °C in 26.7% glycerol (w/v). An LB agar plate was used to culture *E. coli* which was inoculated on the plate with a loopful of culture from a thawed glycerol stock. The plate was subsequently incubated at 37 °C for 18-24 hours. A single colony on the plate was used to inoculate 5 mL of medium which was then incubated at 37 °C in a shaker incubator for 18-24 hours for the ultimate production of a broth culture. Two mL of this 5 mL media culture was used to inoculate 200 mL of TB which was incubated at 37 °C in a shaker incubator for 2-4 hours for protein expression. Glycerol stocks were prepared from 1 mL of LB culture with the addition of 500 µL of an 80 % glycerol solution (w/v) following storage at -80 °C. Table 2.1 shows the bacterial strains used and their phenotypes.

3.3 *E. coli* expression cultures

From a working glycerol stock, bacteria were inoculated on a LB agar plate containing the appropriate antibiotics (see section 3.1). A single colony of the bacteria with the expression plasmid of interest was selected to inoculate a sterilin tube containing 5 mL of LB and antibiotic. This culture was subsequently incubated overnight at 37 °C with a stir bar continuously stirring at 200 rpm. A sample of 2 mL of this overnight culture was used to

inoculate a previously autoclaved 1 L baffled Erlenmeyer flask which contained 200 mL of TB media and appropriate antibiotic. The TB media culture was allowed to grow at 37 °C in a shaker incubator at 200 rpm until an A_{600} of 0.4-0.6 was reached as this range is indicative of the mid-late exponential phase. In order to induce the expression of proteins, IPTG was added to give a final concentration level of 100 µg/mL. Finally, the culture was incubated at 30 °C for 18-20 hours followed by centrifugation at 5,000 rpm for 10 minutes using a Sorvall™ GSA rotor for collection of the cells. The pellet was then stored at -20 °C, and the supernatant was autoclaved and discarded accordingly.

3.4 Preparation of cleared lysate for protein purification

3.4.1 Cell lysis by cell disruption

In order to resuspend the cell pellet that resulted from the centrifugation of a 200 mL TB culture, 100 mL of lysis buffer containing 20 mM imidazole and 0.1% antifoam (w/v) was added (Sigma Antifoam SE-15). With the aid of a magnetic bar and stirrer, the cell pellet was mixed in this lysis buffer for 10-15 minutes to fully dissolve and homogenise the cells in the buffer. Cells lysis was achieved using a Constant Systems Ltd cell disruptor (TS Series Bench Top) and the pressure to breakdown the cells was selected according to the organism under disruption. For *E. coli,* the pressure selected was 15 kpsi (kilopound per square inch). With the purpose of equilibrating the disruptor, 100 mL of lysis buffer containing 20 mM imidazole was passed through the system. As proteins require lower temperatures for the preservation of their biological functions, the disruptor was kept cooled by preparing a water bath filled with ice water and circulating this through the cooling jacket which surrounds the disruption head. The resuspended cell sample was then loaded into the reservoir where the cellular disruption took place. The sample was then loaded into the machine's reservoir for a second round for assurance of complete cell disruption. An additional 20 mL of lysis buffer was passed through the machine at the selected pressure to capture any residual sample. The system was thoroughly cleaned after each use by running 250 mL of distilled water first, then 250 mL of ethanol/IMS and lastly, 250 mL of distilled water was run through the machine. To collect any

insoluble debris, the disrupted sample was then spun at 10,000 rpm for 40 minutes at 4 °C. The cell lysate supernatant was stored at 4 °C for subsequent processing (protein purification).

3.4.2 Preparation of lysate for IMAC column loading

The preparation of the cell lysate involved the filtration through a Whatman® filter paper (Grade 1 – 11 µm) using a Nalgene® reusable vacuum filter unit (DS0320-5045) into a clean Duran bottle.

3.4.3 Standard IMAC procedure

IMAC-Sepharose resin (GE Healthcare) and Profinity™ IMAC resin (Bio-rad) were the resins used to purify the his-tagged recombinant protein. Two to five mL of resin was loaded into a 20 mL column. The required storage buffer, 20 % (v/v) industrial methylated spirits (IMS), was passed through the column to wash the resin by using 5-10 column volumes (CV) of dH_2O to the point at which the resin was fully settled. Following that, the column was equilibrated with 10 CV of lysis buffer containing 20 mM of imidazole. Subsequently, the filtered cell lysate was loaded into the column and a slow flow rate was set to increase the chances of the occurrence of optimal capture of the protein of interest by the resin. In order to wash the resin, 10 CV of lysis buffer containing 20 mM of imidazole was passed through the column and further wash steps were performed with 5-10 CV of lysis buffer containing 50-80 mM of imidazole. The recombinant His-tagged protein was then eluted using 12-15 mL of a lysis buffer containing a high concentration of imidazole (250 mM). The initial flow through (unbound), washes and elution fraction were all collected and labelled. The column was then washed with 10 CV of distilled water followed by 5 CV of 20 % (v/v) ethanol in which the resin was then stored. Fractions taken at each step of the process were analysed by SDS-PAGE.

3.4.4 Stripping and Recharging the IMAC Resin

Prior to loading a filtered cell lysate, the resin was stripped and recharged. Firstly, the column was washed with 2 CV of dH_2O followed by 2 CV of 50 % IMS (v/v). By loading the column with 2 CV of 100 mM EDTA at pH 8.0, the metal ions were then stripped. In order to remove any remaining impurities, the column was washed with 2 CV of 200 mM NaCl, 2 CV of dH_2O and 10 CV of 30 % isopropanol (v/v). The resin was then washed with 10 CV of dH_2O and recharged with 1 CV of 100 mM $NiSO_4$. Once again, the column was washed with 10 CV of dH_2O and stored in 20 % IMS (v/v).

3.5 Protein quantification using the BCA assay

Total protein was quantified with the aid of the Pierce™ BCA Protein Assay Kit (Thermo Scientific 23227). The kit was used according to the manufacturer's instructions to quantify total protein concentration level with a working range of 20-2,000 µg/mL. The BCA working reagent (WR) was prepared by combining reagent A with reagent B (50:1) which were both supplied in the kit. A 96-well microplate was used to perform the assay in which 200 µL of WR was added to 25 µL of protein sample or BSA standard and repeated in triplicate. The plate was then placed in a shaker for 30 seconds in order to mix the contents and subsequently incubated at 37 °C for 30 minutes. The plate was allowed to cool down to room temperature and the absorbance was read at 570 nm (within the suitable range of 540-590 nm). The data collected allowed the creation of a standard curve and its second order polynomial trendline equation was used to mathematically determine the protein concentration level of the samples.

3.6 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were analysed using the sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) method which is outlined by Laemmli (1970).

3.6.1 Preparation of SDS gels

The gels were made according to Table 3.1.

Table 3.1: SDS-PAGE gel recipes

Components	15 % Separating gel	4 % Stacking gel
Acrylamide/Bis-acrylamide, 30% solution	3.750 mL	325 µL
dH_2O (w/v)	1.758 mL	1.54 mL
1.5 M Tris-HCl pH 8.8	1.875 mL	-
0.5 M Tris-HCl pH 6.8	-	625 µL
10 % (w/v) Ammonium persulphate (APS)	37.5 µL	12.5 µL
10 % (w/v) SDS	75 µL	25 µL
TEMED (*N,N,N′,N′*-tetramethylethane-1,2-diamine)	3.75 µL	2.5 µL

As APS and TEMED are polymerizing agents, they were added last to the mixes (both to the separating and stacking gels). Gels were cast using ATTO mini slab glass plates of dimensions 90 x 80 x 1 mm. The glass plates and gasket were assembled and held with clips and the seal was checked by pouring 70 % IMS (v/v) prior to loading the gel. Shortly after the addition of APS and TEMED to the mix, the separating gel was poured to about 1.5 cm below the top of the plate. The separating gel was overlaid with 70 % IMS (v/v) and allowed to polymerise at room temperature for 45-50 min. The importance of overlaying the separating gel relies on the fact that the 70 % IMS solution ensures that the top of the gel is completely flat. This keeps

protein sample parallel to the bottom of the wells and prevents the separating gel from dehydration after it polymerises. The 70 % IMS layer is removed after the fully polymerisation of the gel and the stacking gel is added. A comb was then inserted diagonally, to ensure air bubbles did not remain at the bottom of the well, into to the top of the stacking gel to form the wells in which samples can be loaded into. The gel was allowed to polymerise at room temperature for 30 minutes. In the case where the gel was not immediately used, it was wrapped in damp tissue and stored at 4 °C.

3.6.2 Protein sample preparation and application

Twenty microliters of sample and 4 µL of SDS Sample Buffer, 6X (see section 2.5) were added to a 1.5 mL microcentrifuge tube. This mixture was then boiled at 100 °C on a heating bloc (Labnet Accublock™ Digital Dry Bath) for 5 minutes. The comb and gasket were removed and the gel (sandwiched in the two glass plates) was placed in the electrophoresis chamber. A certain volume of 1X SDS-PAGE running buffer was added to the chamber allowing the removal of unpolymerised acrylamide from the wells. Twenty µL of the prepared sample was applied to each SDS-PAGE well. The first lane of the gel was used to load 15 µL of the molecular weight marker from Thermo Scientific (PageRuler™ Plus Prestained Protein Ladder of code 26619, Figure 3.1). Gel was run at 25 mA for 30 minutes and then at 45 mA for about 50 minutes. The ATTO AE-6500 mini-slab size electrophoresis system was used connected to a Labnet Power Station™ 300 power supply.

The Thermo Scientific PageRuler™ Plus Prestained Protein Ladder (Figure 3.1) consists of a mixture of nine recombinant proteins ranging from 10 KDa to 250 KDa. There are two orange bands at 70 KDa and 25 KDa and a green reference band at 10 KDa to highlight the protein ladder. The remaining six bands are stained blue.

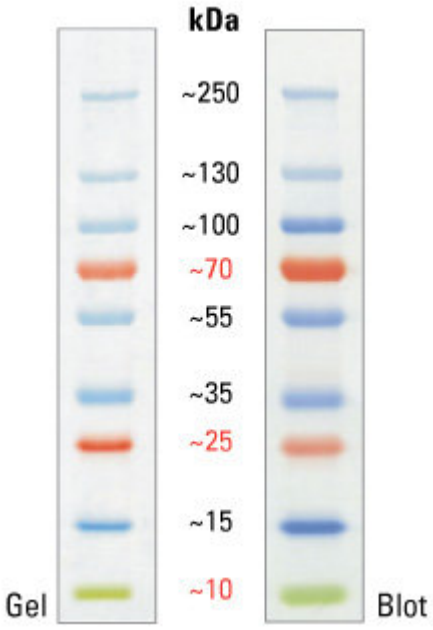

Figure 3.1: SDS-PAGE band profile of the PageRuler Plus Prestained Protein Ladder (26619). Images are from a 4-20% (w/v) Tris-glycine gel (SDS-PAGE) and subsequent transfer to membrane. Images are from www.thermofisher.com.

3.6.3 Staining SDS-PAGE gels

After carefully removing the polyacrylamide gels from between the glass plates using a spatula and rinsing them with dH_2O, the gels were left to stain for a minimum of 2 hours with Coomassie blue stain solution (see section 2.5) in a plastic weigh boat. The boat was placed on an orbital shaker set at a slow rotation at room temperature. The stain solution was removed, and the gels were rinsed with dH_2O again before the addition of Coomassie blue destain solution. The gels were left on the destain solution for 2 hours and additional Coomassie blue destain solution was added as required until the proteins bands were visible, and the gels were free from the blue background; in other words, gels were transparent. Once the gels were fully destained, they were rinsed with dH_2O and placed in a clean weigh boat and a digital camera was used to obtain gel images which were captured on a F1.9 16 MP Smart OIS camera of a Samsung Galaxy S6 phone.

3.7 Buffer exchange of recombinant protein fractions

As a result of the IMAC protein purification process steps, the expressed protein is eluted into lysis buffer containing 250 mM imidazole. In order to accurately quantify the amount of protein, keep it biologically active and store in an optimal buffer, it is necessary to exchange the buffer. The elution fractions to be purified were pooled and passed through a spin column with a molecular weight cut-off (MWCO) of 10 KDa. The MWCO selected should be at least 50 % smaller than the protein of interest. For volumes greater than 5 mL, the Vivaspin® Turbo 15 from Sartorius (VS15T02 – max speed 4,000 x g) was used. However, for volumes less than 5 mL, the Vivaspin® 500 (VS0102 – max speed 15,000 x g) was used. The suitable spin column was loaded with the pooled elution fractions and centrifuged at maximum speed for 10 minutes at room temperature. The flow through was collected and the retentate was topped up with the preferred buffer and the centrifuge step was repeated a further 3-5 times. Following that, the protein was suitable for quantification analysis or storage at -20 °C (short term) or -80 °C (long term).

3.8 Biotinylation of recombinant proteins

The biotinylation of recombinant proteins is a required process to enable the conduction of activity assays and to probe live cells using a flow cytometer. The Thermo Scientific EZ-Link™ Sulfo-NHS-SS-Biotin No-Weigh ™ Format (21328) was used to biotinylate recombinant proteins. The biotinylation process relies on the reaction of *N*-Hydroxysuccinimide (NHS) activated biotins with primary amino groups, $-NH_2$, in pH 7-9 buffers for the formation of stable amide bonds. Usually, proteins have several primary amines in the side chains of lysine (K) residues and at the N-terminus of the polypeptide which are available for labelling with NHS-activated biotin. Biotin (B_7 vitamin) binds with high affinity to avidin and streptavidin and has useful multiple features such as solubility in water, small size molecule (244 Da) and non-interference in the biological protein activity when biotin is conjugated to the protein.

The kit was used according to the manufacturer's instructions. Therefore, the kit was removed from the freezer and a 10 mM biotin solution was prepared by adding 164 µL of ultrapure water to 1 mg of biotin in a microtube. In order to calculate the amount in milimoles of Sulfo-NHS-SS-Biotin to add to the reaction for a 20-fold molar excess solution, Equation 3.1 (below) was used. Then, Equation 3.2 was used to calculate the amount in microliters of the 10 mM Sulfo-NHS-SS-Biotin to add to the reaction.

$$mL\ protein\ x\ \frac{mg\ protein}{mL\ protein}\ x\ \frac{mmol\ protein}{mg\ protein}\ x\ \frac{20\ mmol\ Biotin}{mmol\ protein} = mmol\ Biotin \qquad \text{(Equation 3.1)}$$

$$mmol\ Biotin\ x\ \frac{607\ mg}{mmol\ Biotin}\ x\ \frac{1000\ \mu L}{6.0\ mg} = \mu L\ Biotin \qquad \text{(Equation 3.2)}$$

Legend:

- 20 = Recommended molar fold excess of biotin for 2 mg/mL IgG sample
- 607 = Molecular weight of Sulfo-NHS-SS-Biotin
- 1000 = Microliters of water in which 6.0 mg of Sulfo-NHS-SS-Biotin is dissolved to yield a 10 mM solution

An amine free buffer such as PBS was used for the biotinylation of proteins. The calculated volume of biotin was added to the protein and incubated for 2 hours or at room temperature for 30 minutes. The sample was then buffer exchanged (see section 3.7), for biotin removal and increase of the protein concentration, using a 10,000 Da MWCO spin column as described in section 3.7.

3.9 Enzyme-linked lectin assay

McCoy et al. (1983) was the first to describe the enzyme-linked lectin assay (ELLA) for detecting glycoproteins bearing specific carbohydrate residues. In the present research work, the method described and optimised outlined by Thompson et al. (2011) was used for the characterization of lectins and the determination of the lectin binding specificities (Figure 3.2).

A 50 µL volume of glycoprotein at 5 µg/mL was immobilized in each well of a Nunc-Immuno™ MicroWell™ 96 well solid plate (439454) and incubated at 4 °C for 16-18 hours. Each sample was set up in triplicate. By inverting the plates, the unbound glycoprotein was removed and following that the wells were blocked with 150 µL of 0.5 % (w/v) polyvinyl alcohol (PVA) in TBS for one hour at 25 °C. After inverting the plate to remove the blocking solution, the plate was washed with TBS supplemented with 0.1 % (v/v) Tween 20 four times. A 50 µL aliquot of lectin in TBS supplemented with 1 mM $CaCl_2$ was then added at a concentration of 5 µg/mL and incubated at 25 °C for 1 hour. This solution was then removed by inverting the plate and washed with TBST as previously. This next step was the addition of 50 µL of 1:10,000 murine anti-histidine (anti-polyHistidine mAb conjugated to HRP, Sigma A7058) and/or anti-biotin antibody (anti-Biotin mAb conjugated to HRP, Sigma A0185), as appropriate. The antibody was diluted fresh in TBST and was incubated for 1 hour at 25 °C. A 100 µL volume of TMB substrate (see section 2.5) was added after the removal of the unbound antibody by inverting the plate and washing it four times with TBST. A volume of 50 µL of 10 % (v/v) H_2SO_4 was used to cease the reaction after a specified time. The microplate was then ready for absorbance reading at 450 nm using a BioTek ELx808 plate reader.

All commercial lectins which have been used in this present research work were supplied by Vector Laboratories Ltd UK (Table 3.2). In order to demonstrate lectin specificity a negative control was used. Each lectin was diluted to 5 µg/mL in TBS supplemented with 10 mM $CaCl_2$ and a defined sugar was added. The used concentration level of the inhibiting/eluting sugar is recommended by Vector Laboratories Ltd in the product data sheet. This lectin-sugar solution was then added to the plate as described above.

Additionally, in-house produced lectins were also used: LEC A, LEC B and AAL-2. These lectins were developed by Jonathan Cawley and Donal Monaghan during their PhD work in the research group. LEC A specifically binds to Galactose, LEC B to Fucose and Mannose, and AAL-2 to *N*-Acetylglucosamine.

Table 3.2: Biotinylated lectins from Vector Laboratories Ltd UK. (Man = Mannose; GlcNAc = *N*-Acetylglucosamine; Gal = Galactose; Lac = Lactose; SA = Sialic Acid; Fuc = Fucose; Glu = Glucose and GalNAc = *N*-Acetylgalactosamine)

Lectin Name (Abbreviation)	**Lectin Source**	**Common Name**	**Binding Specificity**
Concanavalin A (CONA)	*Canavalia ensiformis*	Jack Bean	Man, Glu
Ricinus communis agglutinin (RCA)	*Ricinus communis*	Castor oil plant	GlcNAc, Gal
Aleuria aurantia Lectin (AAL)	*Aleuria aurantia*	Orange Peel Fungus	Fuc
Maackia amurensis Lectin II (MAL II)	*Amur maackia*	Amur tree	SA
Wheat germ agglutinin (WGA)	Wheat	*Triticum* spp.	GlcNAc
Peanut agglutinin (PNA)	*Arachis hypogaea* peanuts	Peanut	Gal

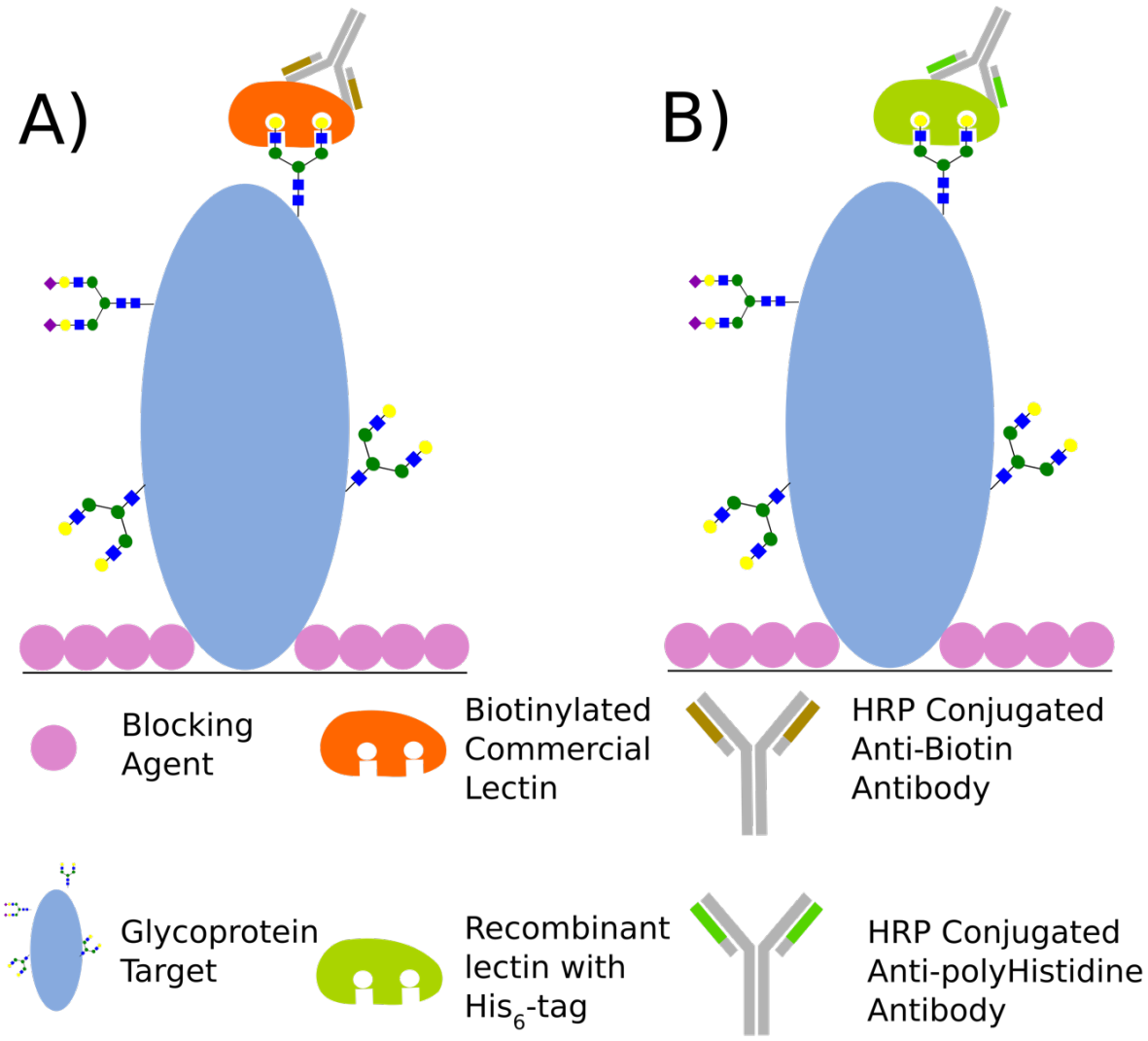

Figure 3.2: Schematic diagram of an ELLA. An immuno microplate is blocked with 0.5 % (w/v) PVA in TBS after a glycoprotein target is immobilized on the surface of the plate. Lectins are added to interact with the glycoproteins and then antibodies are added for bound lectin detection. A) Glycoproteins probed with biotinylated commercial lectins. B) Glycoproteins probed with His_6-tag recombinant lectins. The image was generated with the aid of Inkscape 0.91.

3.10 Mycoplasma Testing

Mycoplasma contamination is of a permanent concern in animal cell culture, so in order to check cells and expand the CHO-K1 cell stock (see section 3.11.5), mycoplasma testing was done using the MycoAlert™ Plus detection kit from Lonza. The assay was performed on a Glomax™ luminometer and conducted according to the protocol suggested by the manufacturer.

3.11 Cell culture techniques

3.11.1 General consumables

The sterile plastic consumables used for cell culture in this present work, such as 96 well plates, 50 mL tubes, pipette tips (10 µL – 25 mL) and microcentrifuge tubes (not supplied sterile, so tubes were sterilized by autoclaving at 121 °C for 20 minutes), were mostly supplied by Sarstedt. Sterile 20 mL tubes were supplied by Thermo Scientific. Sterile 50 mL CELLSTAR®CELLreactor™ TUBES were used to culture cells in suspension and the tubes were supplied by Greiner Bio-One.

3.11.2 Cell culture cabinet and incubators

All cell culture work was conducted in a Holten Laminar HB2448 cabinet and a HERAsafe KS18 class II biological safety cabinet (BSC). Aseptic techniques were used at all times to ensure cells were protected from contamination. The BSC was thoroughly sprayed and wiped down with 70 % (v/v) IMS before and after use. All the items which needed to be manipulated in the cabinet were sprayed with 70 % IMS. In order to diminish the risk of cross-contamination, only one cell line was manipulated at a time and the BSC was left vacant for a minimum of 15 minutes between different cell lines. An ORBi-SHAKER™ CO_2 19mm orbital shaker from

Benchmark was placed inside of a Heraeus® Function Line CO_2 incubator, and an Advanced Mini Shaker 15 mm orbital shaker from VWR was placed inside of a Memmert CO_2 incubator INB200 for cell culture in suspension. Both, incubators and BSC, were regularly cleaned with a broad spectrum disinfectant Virkon® (1 % w/v) followed by distilled water and IMS.

3.11.3 Subculture of CHO-K1

CHO-K1 cells were maintained at 37 °C in an atmosphere with 5 % CO_2 and about 95 % humidity. Cell were grown in 5 mL of medium in 50 mL bioreactor tubes and placed on an orbital shaker (Heraeus Function Line CO_2 Incubator BB 16) which was set at 200 rpm. Subculture of CHO-K1 cells was conducted every 3 to 4 days and cells were only used for experiments up to a maximum of 10 passages after recovery (see section 3.11.6).

The bioreactor tube was spun at 1000 rpm for 5 min to collect the cells. After carefully removing the supernatant, cells were resuspended in 5 mL of fresh pre-warmed growth medium (see section 2.4). Cell counting was then counted on a haemocytometer following staining with Trypan Blue solution; an appropriate volume of the cell suspension was then used to seed fresh tubes at a desired starting cell density.

3.11.4 Trypan blue cell counts

Cells were counted using an Improved Neubauer haemocytometer (Hawksley BS.748). In order to determine cell viability, Trypan Blue solution (0.4 % Trypan Blue, Sigma T8154) was used. A volume of 20 µL of trypan blue solution was added to a sample of 180 µL of cell suspension, mixed and allowed to rest for 5 minutes. A clean glass coverslip (24 mm L x 26 mm W x 0.4 mm H) was moistened with exhaled breath and then the coverslip was slid over the chamber back and forth using slight pressure until Newton's refraction rings appeared. These rings are seen as rainbow-like ones under the coverslip. A volume of 10 µL of the mix

(cell suspension + trypan blue solution) was used to fill one side of the chamber. An Olympus CK40 inverted microscope was used to look at the cells. As trypan blue only enters in non-viable cells, cells stained blue were counted as non-viable whereas the bright cells were counted as viable. The concentration of viable and non-viable cells and the percentage of viable cells were calculated as follows:

$$Viable\ Cell\ Count\ = \frac{No\ of\ Live\ cells\ counted}{\substack{No\ of\ large\ corner \\ Squares\ counted}}\ x\ 10\ (DF)\ x\ 10{,}000 \qquad \text{(Equation 3.3)}$$

$$Non-viable\ Cell\ Count\ = \frac{No\ of\ Dead\ cells\ counted}{\substack{No\ of\ large\ corner \\ Squares\ counted}}\ x\ 10\ (DF)\ x\ 10{,}000 \qquad \text{(Equation 3.4)}$$

$$Percentage\ Viability = \frac{No\ of\ Viable\ Cells}{Total\ No\ of\ Cells}\ x\ 100\% \qquad \text{(Equation 3.5)}$$

The concentration of cells (viable and non-viable) is in cells/mL and the dilution factor (DF) is 10.

3.11.5 Cryopreservation of Cells

For cryopreservation of the cells for indefinite time, cells were stored below -180 °C in a liquid nitrogen tank and an appropriate freezing medium from Gibco® was used (Recovery™ Cell Culture Freezing Medium, Bio-Sciences Ltd 12648010). In order to obtain optimum results, cells were sampled in mid-log phase of growth (3 to 4 days in culture) with > 90 % viability at the time of freezing. The freezing medium was allowed to thaw at 2-8 °C and mixed before use. Cells were counted according to the method described in the previous section 3.11.4 and the required volume of the freezing medium was calculated to achieve a final cell density of 1 x 10^6 to 1 x 10^7 cells/mL. The suspension cells were transferred to a sterile 20 mL centrifuge tube and centrifuged at 1000 rpm for 5 minutes. The supernatant was aseptically removed and cells were resuspended in the required volume of freezing medium. Cells were subsequently dispensed into 1.5 mL cryovials (Sarstedt 72.694.406). The cell suspension was

frequently mixed to ensure homogeneous aliquots were being taken. Cryopreservation was achieved using a manual controlled rate freezing apparatus (Mr. Frosty™ freezing container, Thermo Scientific 5100-0001) which allowed the cells to freeze approximately 1 °C decrease per minute in a -80 °C freezer overnight. The following day, the cryovials were transferred to the liquid nitrogen tank for indefinite storage.

3.11.6 Recovery of cells

In order to recover cells from cryo-storage, they were removed from the liquid nitrogen tank and rapidly thawed (< 1 minute) at 37 °C until only a small amount of ice remained. Cells were then transferred to a sterile 20 mL centrifuge tube and a volume of 5 mL of pre-warmed medium was added. The tube was place on a centrifuge and spun at 1000 rpm for 5 minutes. The supernatant was aseptically removed, and cells were resuspended in 5 mL of pre-warmed medium and transferred to a bioreactor tube for culturing in the incubator.

3.12 Flow cytometry methods and statistical analysis

3.12.1 Sample preparation

Prior to the analysis of cell surface glycosylation, a number of optimization studies was conducted. These studies set out to determine the optimal conditions of certain variables thus allowing an increase in the quality level of the information extracted from the main study.

Additionally, several decisions were made towards the minimization of sources of variabilities which also contributed to the increase of the level of quality of the experimental data. The choice of using industrially manufactured PBS and the BALANCD CHO media to manipulate the cells during sample preparation allowed consistency of the solutions used on the cells. Furthermore, the biotinylated lectins used were from the same batches throughout the entire project as it is known that biotinylated lectins can present batch-to-batch variability with

regard to the number of biotin groups attached to the lectin molecule. The number of biotin groups can, in turn, influence the level of fluorescence intensity as the streptavidin V450 molecules will attach to biotin groups proportionally. Therefore, an experiment might present a glycoprofile variation in relation to a baseline experiment due to the use of a biotinylated lectin from a different batch.

3.12.1.1 Cell culture process optimisation

Two sets of six cell culture tubes were prepared to collect pH measurements and flow cytometry data to construct cell culture growth curves. Each set was composed of two subsets which were distinguished by L-Glutamine concentration levels of 2 mM and 4 mM. Each subset was composed of three tubes seeded at a different starting cell density: 0.5, 1 and 2 million cells/mL.

While a set was used for daily sampling of flow cytometric analysis and pH measurements, the other was only used to collect samples for pH measurements. The sets were labelled as **FC/pH** and **pH** respectively (Figure 3.3).

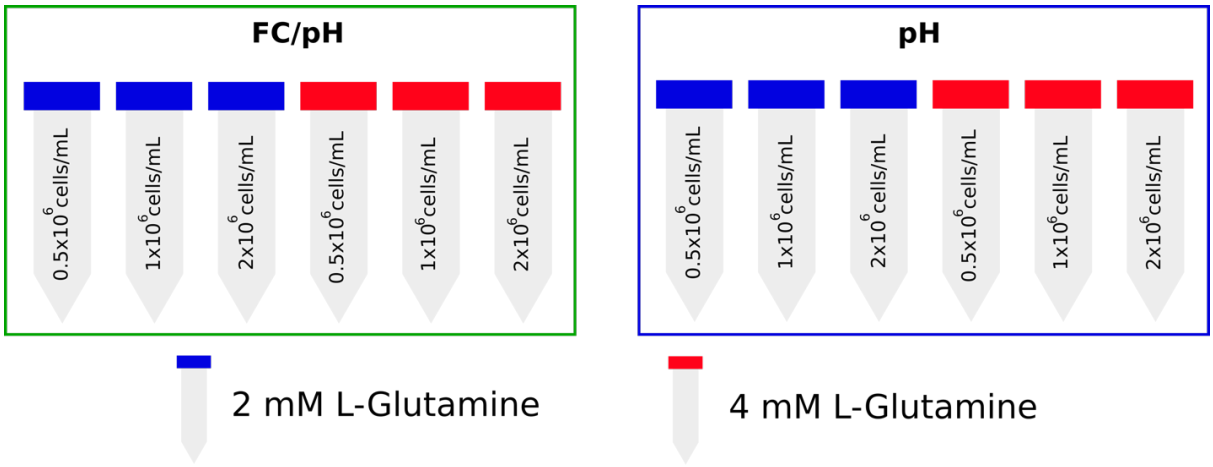

Figure 3.3: Diagram illustrating the arrangement of cell culture tubes for cell culture optimisation studies.

Samples were collected every 24 hours for 10 days. A volume of 200 µL of cell suspension from each tube of the **pH** set was sampled whereas 350 µL from the **FC/pH** tubes. Then, 50 µL of CountBright™ absolute counting beads (ThermoFisher, C36950) was added to each **FC/pH** sample. Also, a 1 µL of 7-AAD was added to the samples followed by an incubation period of 15 minutes at room temperature in the dark.

The **FC/pH** samples were then transferred to FACS tubes and data was collected using a flow cytometer BD FACSAria I. Upon completion of flow cytometry data collection, pH measurements of the **pH** and **FC/pH** samples were taken using a pH electrode Orion Semi-micro (ThermoFisher, 10237293) attached to a Eutech pH510 bench pH meter. Three pH measurements were taken from each sample.

3.12.1.2 7-AAD concentration optimisation

At the fourth day of culture, CHO-K1 cells were counted using the method described in section 3.11.4. A volume of cell suspension containing $2x10^6$ cells was calculated and sampled into microcentrifuge tubes. One milliliter of room temperature sterile PBS (Sigma, D8537) was added to the microcentrifuge tubes. These tubes were then centrifuged at 400 g for 5 minutes and the supernatants obtained were then discarded. Each tube was labelled with the 7-AAD volume to be added to it and a randomly assigned ordinal number. For example, *1 µL of 7-AAD – 2.* Five different concentration levels of 7-AAD were tested by adding the following volumes: 1, 2, 3, 4, and 5 µL. Cells were resuspended in 500 µL of supplemented pre-warmed medium (at 37 °C for an hour) and the respective amount of 7-AAD was added. Cells were incubated for 15 minutes in the dark at room temperature. A tube of unstained cells was also prepared in parallel to act as a flow cytometry control. The tube contents were then transferred into labelled FACS tubes which were then kept in a styrofoam box full of small ice cubes to reduce cell metabolism. Shortly after, 10,000 events of the singlets population (see section 3.12.5) were collected from each sample starting with the flow cytometry control sample.

3.12.1.3 7-AAD incubation time optimisation

At the fourth day of culture, CHO-K1 cells were counted using the method described in section 3.11.4. A volume of cell suspension containing $2x10^6$ cells was calculated and sampled into microcentrifuge tubes. One milliliter of room temperature sterile PBS (Sigma, D8537) was added to the microcentrifuge tubes. These tubes were then centrifuged at 400 g for 5 minutes and the supernatants obtained were then discarded. Three different 7-AAD incubation time periods were tested: 5, 10 and 15 minutes. Cells were resuspended in 500 µL of supplemented pre-warmed medium (at 37 °C for an hour) and 1 µL of 7-AAD was added to the 5 min samples first. These samples were incubated for 5 minutes in the dark at room temperature. A tube of unstained cells was also prepared in parallel to act as a flow cytometry control. The tube content was then transferred into a labelled FACS tube which was then kept in a styrofoam box full of small ice cubes to reduce the cell metabolism. Shortly after, 10,000 events of the singlets population (see section 3.12.5) were collected from each sample starting with the flow cytometry control sample.

Upon the conclusion of the reading of the 5 min sample, 1 µL of 7-AAD was added to the 10 min samples and the incubation time was conducted in the dark at room temperature. The same procedure was then followed as previously. Finally, the 15 min sample set was then prepared by adding 1 µL of 7-AAD and 15 minutes of incubation time was conducted in the dark at room temperature. Again, the same procedure was followed as previously concluding the experiment.

3.12.1.4 DRAQ5 concentration optimisation

At the fourth day of culture, CHO-K1 cells were counted using the method described in section 3.11.4. A volume of cell suspension containing $2x10^6$ cells was calculated and sampled into microcentrifuge tubes. One milliliter of room temperature sterile PBS (Sigma, D8537) was added to the microcentrifuge tubes. These tubes were then centrifuged at 400 g for 5 minutes

and the supernatants obtained were then discarded. Each tube was labelled with the DRAQ5 volume to be added to it and a randomly assigned ordinal number. For example, *1 µL of DRAQ5 – 3.* Five different concentration levels of DRAQ5 were tested by adding the following volumes: 1, 2, 3, 4, and 5 µL. Cell were resuspended in 500 µL of supplemented pre-warmed medium (at 37 °C for an hour) and the respective amount of DRAQ5 was added. Cells were then incubated for 20 minutes at 37 °C. A tube of unstained cells was also prepared in parallel to act as a flow cytometry control. The tube contents were then transferred into labelled FACS tubes which were then kept in a styrofoam box full of small ice cubes to reduce the metabolism of the cells. Shortly after, 10,000 events of the singlets population (see section 3.12.5) were collected from each sample starting with the flow cytometry control sample.

3.12.1.5 DRAQ5 incubation time optimisation

At the fourth day of culture, CHO-K1 cells were counted using the method described in section 3.11.4. A volume of cell suspension containing $2x10^6$ cells was calculated and sampled into microcentrifuge tubes. One milliter of room temperature sterile PBS (Sigma, D8537) was added to the microcentrifuge tubes. These tubes were then centrifuged at 400 g for 5 minutes. The supernatants obtained were then discarded. Five different DRAQ5 incubation time periods were tested: 5, 10, 15, 20 and 25 minutes. Cell were resuspended in 500 µL of supplemented pre-warmed medium (at 37 °C for an hour) and 1 µL of DRAQ5 was added to the 5 min samples first. These samples were incubated for 5 minutes at 37 °C in the incubator. Meanwhile, the 10 min sample was prepared and allowed to incubate for the expected time and a tube with unstained cells was also prepared in parallel to act as a flow cytometry control. After the 5 min sample completed the incubation step, the tube content was then transferred into a properly labelled FACS tube which was then kept in a Styrofoam box full of small ice cubes to reduce cell metabolism. Shortly after, 10,000 events of the singlets population (see section 3.12.5) were collected from each sample starting with the flow cytometry control sample.

After the completion of the data collection of the 5 min sample, the 15 min sample was prepared and allowed to incubate as expected. As soon as the 10 min sample completed the incubation time, this sample was transferred to a FACS tube and read. Then the 20 min sample was prepared and incubated for the expected time. Meanwhile, the 15 min sample was ready to be transferred to a FACS tube for data collection. Finally, the 25 min sample was prepared and incubated. At this point, the 20 min sample was transferred to a FACS tube and read. Upon completion of the expected incubation time, the 25 min sample was prepared for data collection and the last experimental reading was performed.

3.12.1.6 Lectin Cytotoxicity Analysis

At the fourth day of culture, CHO-K1 cells were counted using the method described in section 3.11.4. A volume of cell suspension containing $2x10^6$ cells was calculated and sampled into microcentrifuge tubes. One milliter of room temperature sterile PBS (Sigma, D8537) was added to the microcentrifuge tubes which was followed a centrifugation step at 400 g for 5 minutes.

Meanwhile, serial dilutions of individual lectins were prepared starting at 12.50 µg/mL of lectin and 6.25 µg/mL of V450 concentration levels in a volume of 1.4 mL of supplemented pre-warmed medium (at 37 °C for an hour). The serial dilutions resulted in 8 lectin concentration levels: 12.50, 6.25, 3.13, 1.56, 0.78, 0.39, 0.20, and 0.10 µg/mL. Samples at 0.00 µg/mL of lectin/V450 were also prepared. Each tube was labelled with its replicate number, lectin concentration level, and a randomly assigned ordinal number, for example, *Replicate I - WGA at 12.5 µg/mL - 12*. The ordinal number assigned to each tube was previously generated using the *sample( )* function on R which can randomly order a sequence of numbers. The ordinal number dictated the order of the treatment assigned for each tube as well as every other action in the sample preparation process (except for the centrifugation of the tubes which was conducted at the same time), including the order at the data collection stage. Such way of arranging the order of treatment application and data collection was employed to remove the time as a confounding parameter.

Following centrifugation, supernatants were removed as above and cells were resuspended in 200 µL of the appropriate lectin/V450 solution and incubated in the dark for 40 minutes at room temperature. The tubes were centrifuged again at 400 g for 5 minutes and the supernatants containing the unbound lectin molecules were removed to reduce the level of background signal. Cells were resuspended in 500 µL of supplemented pre-warmed medium (at 37 °C for an hour) and 1 µL of 7-AAD was added. Cells were incubated for 15 minutes in the dark at room temperature.

Another set of microcentrifuge tubes was prepared in parallel to act as flow cytometric control samples. These samples were composed of two single-stained tubes; 7-AAD and the tested lectin. This lectin single-stained tube was prepared at the highest lectin concentration level which was being tested, i.e, 12 µg/mL. This set of tubes was treated in the same way regarding the staining step which was not needed for a particular tube. For instance, although a LEC A single-stained tube does not need to be incubated with 7-AAD, this tube was kept in the dark at room temperature during the 7-AAD incubation time of the double-stained tubes (lectin and 7-AAD stained tubes).

The tube contents were then transferred into labelled FACS tubes which were kept on ice. Shortly after, data was collected from the samples starting with the flow cytometric control samples. Around 500,000 events were collected for each tube from the viable cell population (see section 3.12.5).

3.12.1.7 Free sugar inhibition analysis

At the fourth day of culture, CHO-K1 cells were counted using the method described in see section 3.11.4. A volume of cell suspension containing $2x10^6$ cells was calculated and sampled into microcentrifuge tubes. One mL of room temperature sterile PBS (Sigma, D8537) was added to the microcentrifuge tubes. These tubes were then centrifuged at 400 g for 5 minutes. After the completion of the centrifugation, the supernatant of the tubes was extracted and discarded.

Lectin/V450 solutions (WGA, PNA, MAL II, AAL, AAL-2, LEC A, and LEC B) were prepared at 3.0 µg/mL of lectin and 1.5 µg/mL of V450 in 0.9 mL of fully supplemented pre-warmed media (at 37 °C for an hour) containing a particular free-sugar molecule (Table 3.3). Lectin/V450 solutions in 0.9 mL of fully supplemented pre-warmed media without any sugar and a solution containing only V450 at 1.5 µg/mL in 0.9 mL of fully supplemented pre-warmed media were also prepared.

Table 3.3: Lectins and the respective sugar molecules used to prepare solutions for sugar inhibition studies.

Solution	Lectin	Sugar	Sugar concentration	Sugar supplier and product code
1	WGA	*N*-Acetylglucosamine	0.4M	Sigma – A8625
2	WGA	*N*-Acetylgalactosamine	0.4M	Sigma – A2795
3	AAL-2	*N*-Acetylglucosamine	100mM	Sigma – A8625
4	AAL-2	*N*-Acetylgalactosamine	100mM	Sigma – A2795
5	AAL	*L*-Fucose	0.2M	Sigma – F2252
6	AAL	Mannose	0.2M	Sigma - 63582
7	MAL II	Sialic Acid	0.2M	VectorLabs – S9008
8	MAL II	*N*-Acetylglucosamine	0.2M	Sigma – A8625
9	LEC A	Galactose	0.1M	Sigma - 15522
10	LEC A	Mannose	0.1M	Sigma - 63582
11	LEC B	*L*-Fucose	0.2M	Sigma – F2252
12	LEC B	Mannose	0.2M	Sigma - 63582
13	LEC B	Galactose	0.2M	Sigma - 15522
14	LEC B	*N*-Acetylglucosamine	0.2M	Sigma – A8625
15	PNA	Galactose	0.1M	Sigma - 15522
16	PNA	Mannose	0.1M	Sigma - 63582

Each tube was labelled with its replicate number, solution of treatment and a randomly assigned ordinal number, for example, *Replicate III – LEC A + Galactose - 9*. The ordinal number assigned to each tube was generated as described in section 3.12.1.6.

After the labelling process, cells were then re-suspended in 200 µL of the appropriate solution and incubated for 40 minutes in the dark at room temperature. Cells were collected by centrifugation again at 400 g for 5 minutes and the supernatant containing the unbound molecules was removed to reduce the level of background signal. Cell were resuspended in 500 µL of supplemented pre-warmed medium (at 37 °C for an hour). Alongside, tubes of unstained cells were also prepared to act as flow cytometric controls. The tube contents were then transferred into properly labelled FACS tubes which were then kept on ice. Shortly after, 10,000 events of the singlets population (see section 3.12.5) were collected from each sample starting with the flow cytometry control sample.

3.12.1.8 Cell surface glycoprofile analysis

At the fourth day of culture, CHO-K1 cells were counted using the method described in see section 3.11.4. A volume of cell suspension containing $2x10^6$ cells was calculated and sampled into microcentrifuge tubes. One milliter of room temperature sterile PBS (Sigma, D8537) was added to the microcentrifuge tubes which were then centrifuged at 400 g for 5 minutes.

Meanwhile, lectin/V450 solutions (WGA, PNA, MAL II, AAL, AAL-2, LEC A, and LEC B) were prepared at 3.0 µg/mL of lectin and 1.5 µg/mL of V450 in 0.9 mL of supplemented pre-warmed medium (at 37 °C for an hour). The supernatants obtained were then discarded. Each tube was labelled with its replicate number, lectin/V450 solution and a randomly assigned ordinal number; for example, *Replicate II – LEC A - 12*. The ordinal number assigned to each tube was generated as described in section 3.12.1.6.

After the centrifugation, removal of the supernatant and tube labelling process as described above, cells in the tubes were then re-suspended in 200 µL of the appropriate lectin/V450 solution and incubated for 40 minutes in the dark at room temperature. Meanwhile, a sample

of 200 µL of each replicate was collected for pH measurement using a pH electrode Orion Semi-micro (ThermoFisher, 10237293) attached to a Eutech pH510 bench pH meter. Three pH measurements were taken from each replicate. Also, the remaining cell suspension in the replicate cell culture tubes was centrifuged at 1000 rpm for 5 minutes to collect the supernatant and cells separately for posterior analysis. The supernatant and cells were stored in a -20 °C freezer.

Following the lectin incubation time, the microcentrifuge tubes were centrifuged again at 400 g for 5 minutes and the supernatant containing the unbound lectin molecules was removed to reduce the level of background signal. Cells were resuspended in 500 µL of supplemented pre-warmed medium (at 37 °C for an hour) and 1 µL of 7-AAD was added. Cells were incubated for 15 minutes in the dark at room temperature. Subsequently, 1 µL of DRAQ5 was added to the tubes then cells were incubated at 37 °C for 25 minutes.

A set of microcentrifuge tubes was prepared in parallel to act as flow cytometric control samples. These samples were composed of one unstained sample and nine single-stained ones (DRAQ5, 7AAD, LEC A, LEC B, AAL, MAL II, AAL-2, PNA, and WGA). The set of tubes were treated in the same way regarding the staining steps which were not needed for a particular tube. For instance, the single-stained LEC A tube was not incubated with 7AAD and DRAQ5. However, this tube was kept in the dark at room temperature during the 7-AAD incubation step and in the incubator at 37 °C during the DRAQ5 incubation step.

The tube contents were then transferred into properly labelled FACS tubes which were then kept on ice. Shortly after, data were collected from the samples starting with the flow cytometric control samples. Around 500,000 events were collected for each tube from the alive cell population (see section 3.12.5).

3.12.2 Experimental setup

3.12.2.1 CHO-K1 cell culture parameters

CHO-K1 cells were cultured under the relevant conditions to meet each experiment objectives. The cell culture optimisation experiment was the first one to be conducted as it provided information on the cell growth curves of cultures with different L-glutamine concentration levels and different starting cell densities. In addition, this experiment provided valuable information on the variation of the pH in the media during the cell culture process and the length of time necessary for cells to achieve the stationary phase. Most importantly, the growth patterns obtained from this study enabled the researcher to identify deviations in the cell behavior throughout the entire research work ensuring cell culture consistency.

The subsequent experiments were then conducted based on the conclusions drawn from the cell culture optimisation studies which established the set point of L-glutamine concentration level and the starting cell density.

Experiments which intended to evaluate outcomes due to the variation of the temperature, CO_2 and nutrient levels, had those parameters varied (one parameter at a time) across a certain range at the third day of the cell culture process while the remaining parameters were still at the set points (Table 3.4). For instance, the nutrient level was changed in the nutrient depletion experiment while the remaining process parameters such as temperature and CO_2 levels were kept at their set points.

Table 3.4: Cell culture conditions adopted in each experiment.

Experiment	CO_2(%)	Temperature (°C)	L-glutamine concentration (mM)	19 mm orbital shaker (rpm)	15 mm orbital shaker (rpm)	Starting cell density (cells/ml)	Spent medium levels (days)
Cell culture Optimisation	5	37	2 and 4	200	-	0.5, 1, and 2x 10^6	-
7AAD Optimisation	5	37	2	200	-	2x10^6	-
DRAQ5 Optimisation	5	37	2	200	-	2x10^6	-
Lectin Cytotoxicity	5	37	2	200	-	2x10^6	-
Sugar Inhibition	5	37	2	200	-	2x10^6	-
Spent medium Variation	5	37	2	200	-	2x10^6	Ranged from -3 to +3
CO_2 Variation	Ranged from 1 to 10	37	2	200	225	2x10^6	-
Temperature Variation	5	Ranged from 32 to 41	2	200	225	2x10^6	-

3.12.2.2 Spent medium variation experimental setup

Cell cultures were subjected to an intervention on the third day of the culture process by replacing the medium of the cells with a spent medium. Cells were then subjected to the spent medium for 24 hours before they were sampled for flow cytometric analysis.

In order to produce different spent medium levels, parallel cell cultures were set up strategically to achieve the desired spent medium level on the third day of culture of the cell

cultures to be interrogated. The levels of spent media were measured in terms of days. For instance, the screening cell cultures subjected to a -2-day spent medium intervention had the original media replaced by media which had been used by parallel cell cultures for 5 days. In other words, the media had been used by the parallel cell cultures for 2 extra days from the third day of culture (intervention day). Therefore, a medium used for 5 days was further depleted in nutrient levels by 2 days in relation to the baseline. Whereas a +2-day spent medium had an excess of nutrient levels in 2 days in relation to the baseline, thus the +2-day medium was used by parallel cells for only a day.

Media replacement was performed by centrifuging both screening and parallel cell cultures at 10,000 rpm for 3 minutes. Then, the supernatant from the screening culture was removed and replaced by the addition of the supernatant from the parallel culture and the screening cells were completely re-suspended by gently pipetting the media down and up twice. The screening cell cultures subjected to a +3-day spent medium intervention had the supernatant replaced by fresh pre-warmed and supplemented media (at 37°C for an hour) instead.

As the centrifugation step was necessary for media replacement, the baseline screening cell culture was also centrifuged at 10,000 rpm for 3 minutes on the third day of culture. The supernatant was then removed and added back to the culture immediately afterwards. Complete resuspension of the screening cells was achieved by gently pipetting the media down and up twice. The basis of this procedure conducted on the baseline samples relies on the removal of the centrifugation step as a confounding variable on the measurement of the experiment outcomes.

3.12.2.3 CO_2 and temperature variation experimental setup

The experiments which involved the variation of CO_2 and temperature levels were conducted using two sets of incubator and shaker, Set A and Set B. Set B (Memmert Incubator Oven INB200 and Advanced Mini Shaker 15 mm orbit-VWR Orbital Shaker) was used to grow cell cultures at the baseline growing conditions, i.e, 5% of CO_2 and at 37°C. Set A (Heraeus

Function Line CO_2 Incubator BB 16 and Benchmark Orbital Shaker Orbi-Shaker™ CO_2 of 19 mm orbit) was used to cultivate the cell cultures under a CO_2 or temperature level variation. Therefore, cultures on the third day of the culture process were removed from Set B and placed on Set A which had the CO_2 or temperature changed to a desired level 30 minutes beforehand. In order to remove the step just described as a confounding variable in the experiment outcomes, baseline cell cultures were also changed from Set B to Set A on the third day of the culture process. However, this time, no change was made in the CO_2 or temperature level.

Additionally, due to a difference in the orbital diameter of the shakers, a new agitation speed was established for Shaker B based on the agitation speed set up for Shaker A. The Set A had been used since the start of this research project and all the previous cell culture work was performed using the shaker of this set at 200 rpm. However, it was then necessary introduce Set B into the experimental work for the continuous growing of cell cultures under the baseline conditions. Therefore, the agitation speed of the shaker in Set B was calculated based on the agitation speed of the shaker in Set A.

The effort of establishing Set B's agitation speed aimed to minimize the variation in the levels of dissolved oxygen in the media as cultures were transferred from a shaker to another of different orbital diameter. Also, this effort prevents dramatic modifications in the growth curve of the cultures since these alterations can insert variabilities in the outcomes of the experiments (Bates, Phillips and O'Bryan, 2011).

The agitation speed of Shaker B was calculated based on Newton's second law of motion which states force = mass x acceleration. By moving a tube from a shaker to a new one with a different orbit diameter, the aim is to determine a new agitation speed which creates the same force driving liquid movement in the tube as on the original shaker. The mass is the same, as the cell culture tubes were transferred from one shaker to the other without any alteration in liquid volume content. The acceleration for each culturing tube is equal to the velocity/radius. As the radius is equal to orbital diameter/2 and velocity equals the agitation speed (in RPM) multiplied by the circumference of the orbit, the new agitation speed for shaker B was then calculated (Bates, Phillips and O'Bryan, 2011).

By performing the mathematical calculations, the following formula was obtained to stablish the new agitation speed (Equation 3.6):

$$r_B = \sqrt{r_A^2 \times \frac{d_A}{d_B}} \qquad \text{(Equation 3.6)}$$

Where:

d_A = the orbital diameter for the original shaker

d_B = the diameter of the new shaker

r_A = the agitation speed in RPM for the original shaker

r_B = the agitation speed in RPM for the new shaker.

Since the d_A is equal to 19 mm, d_B equals to 15 mm and r_A equals to 200 rpm; r_B was determined to be equal to 225 rpm.

3.12.2.4 Technical and biological replicates

All flow cytometric experiments performed in this study collected data using 3 replicates. However, measurements were taken from technical replicates in the sugar inhibition studies and the optimization studies of the 7-AAD and DRAQ. The measurements of the remaining experiments were taken from biological replicates also known as repeats (Table 3.5).

Table 3.5: Type of replicates used in each of flow cytometric experiment

Experiment	Type of replicates
Sugar Inhibition	Technical replicates
7AAD Optimisation	Technical replicates
DRAQ5 Optimisation	Technical replicates
Cell culture Optimisation	Biological replicates
Lectin Cytotoxicity	Biological replicates
Spent medium Variation	Biological replicates
CO_2 Variation	Biological replicates
Temperature Variation	Biological replicates

While technical replicates consist of taking multiple measurements from the same source, biological replicates allow the measurement to be taken from different sources, yet those sources are of the same nature. For instance, the cell culture experiment which involved the variation of the temperature from 37°C to 38 °C on the third of culture was set up 3 times. In other words, 3 cell culture bioreactor tubes were set up at the same time and under the same conditions, then an intervention in the temperature level of these 3 tubes was made on the third day of culture. Therefore, each tube is a replicate which is determined by the same treatment applied; temperature level variation from 37°C to 38°C on the third day of culture.

3.12.3 Calibration and standardization of the BD FACSAria™ I flow cytometer

A method to check the performance of the BD FACSAria™ I machine was employed to minimise the impact of the equipment's natural drifts from day-to-day use on the outcomes of experiments. In addition, an application settings procedure was conducted to ensure the statistical comparability of measurements obtained on different days. For instance, the

experiments in which the cultivating parameters were modified could only be performed over a span of months. A cell culture subjected to a parameter change was in cultivation for the cycle of four days before the flow cytometric analysis which required a full working day (sample preparation and data collection). On the other hand, experiments which could be conducted in one day did not require the creation of application settings as the daily calibration was enough. 7AAD & DRAQ5 optimisation and lectin cytotoxicity studies are examples of experiments that could be performed within a single working day.

The BD digital cytometers are equipped with a fully automated software and reagent system which provides the characterization, setup and tracking (CS&T) of the equipment. Definition and characterization of the baseline performance, optimisation and standardization of the cytometer setup and verification of the cytometer performance are the functions of the CS&T system.

The employment of the CS&T system allows the extraction of consistent and reproducible data every day, the simplification of the design of multicolour experiments, the generation of higher quality data from multicolour experiments, the offset of day-to-day instrument variability, the ability to standardize across experimental runs and instruments, and the early detection of the degradation of cytometer performance.

3.12.3.1 Performance verification

In order to fully characterize the flow cytometer BD FACSAria™ I, CS&T beads (BD™ CS&T beads, product code: 642412, and LOT: 68955) were purchased to be used throughout the entire time period in which the equipment was used for this research work. By following the protocol recommended by the manufacturer, a baseline performance was firstly obtained for the beads purchased and subsequently a performance verification procedure was conducted using the same beads. The latter procedure was then conducted on all experimental days to ensure the cytometer was performing consistently.

3.12.3.2 Creation of application settings

Application settings for the experiments which involved the variation of culturing parameters were created to ensure the standardization of flow cytometry results across experimental runs performed on different days. Based on the baseline report generated by the cytometer's software when the purchased beads were first used on the equipment, application settings were created for the detection channels of interest according to a BD Technical Bulletin (Meinelt *et al.*, 2012). Once the application settings were created and saved on the cytometer system; these application settings could then be applied shortly after the cytometer performance verification was conducted, setting up the equipment for a standardized experimental run on a particular day.

3.12.4 FC panel design and compensation analysis

3.12.4.1 Flow cytometry panel design

As most of the experimental work involved the use of multiple dyes, a careful panel design was developed to provide high quality data within the specifications of the BD FACSAria™ I cytometer. Although V450, 7-AAD and DRAQ5 have overlapping areas in the emission spectrum, 450/40, 610/20 and 780/60 filters, respectively, were selected to allow the minimisation of the spillage of emission signals of the dyes into the areas of neighbouring filters (Figure 3.4).

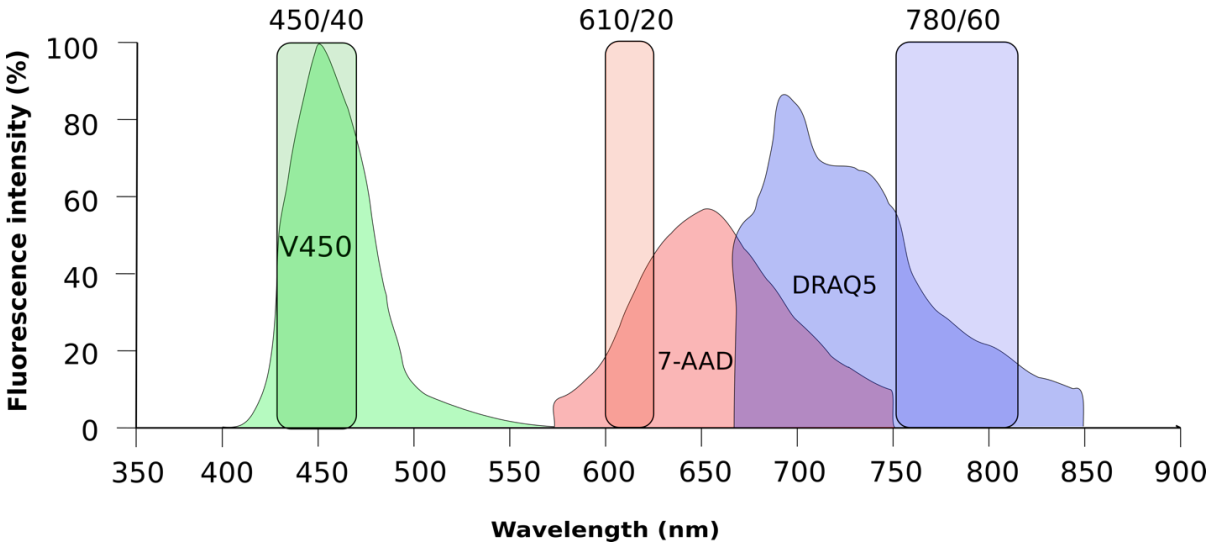

Figure 3.4: Plot illustrating the emission signals from the V450, 7-AAD and DRAQ5 dyes excited at 407 (violet laser), 488 (blue laser) and 633 nm (red laser) respectively. The plot also shows emission overlapping areas and the regions (filters) from which the data was collected. The filters were strategically selected to avoid overlapping emission signals.

3.12.4.2 Compensation analysis

Compensation is a mathematical process which allows the correction of emission signal spillage. Therefore, such a process must be employed prior to the analysis of a multicolour flow cytometry data (Biosciences, 2009). Although the filters selected were in emission regions which do not contain overlapping signals, a statistical analysis was conducted to ensure the non-necessity for data compensation.

The compensation analysis involved the preparation of three single-stained samples (V450/Lectin, 7-AAD, and DRAQ5) and an unstained one (Section 3.12.1). These samples were then interrogated, and data was collected from the LECTIN (V450), 7-AAD and DRAQ5 detector channels for the four samples. Compensation samples were prepared for every experimental run involving the use of more than one fluorescent reagent.

3.12.5 Gating strategies

The acquisition of data from the experiments involving the variation of the three cultivation parameters (spent medium variation, temperature and CO_2 levels) required the use of several detection channels and the application of a set of filters (gates) to classify the cell population into relevant cell subpopulations. The diagram (Figure 3.5) illustrates the sequence in which the different detection channels were used to apply the gates generating the cell subpopulations.

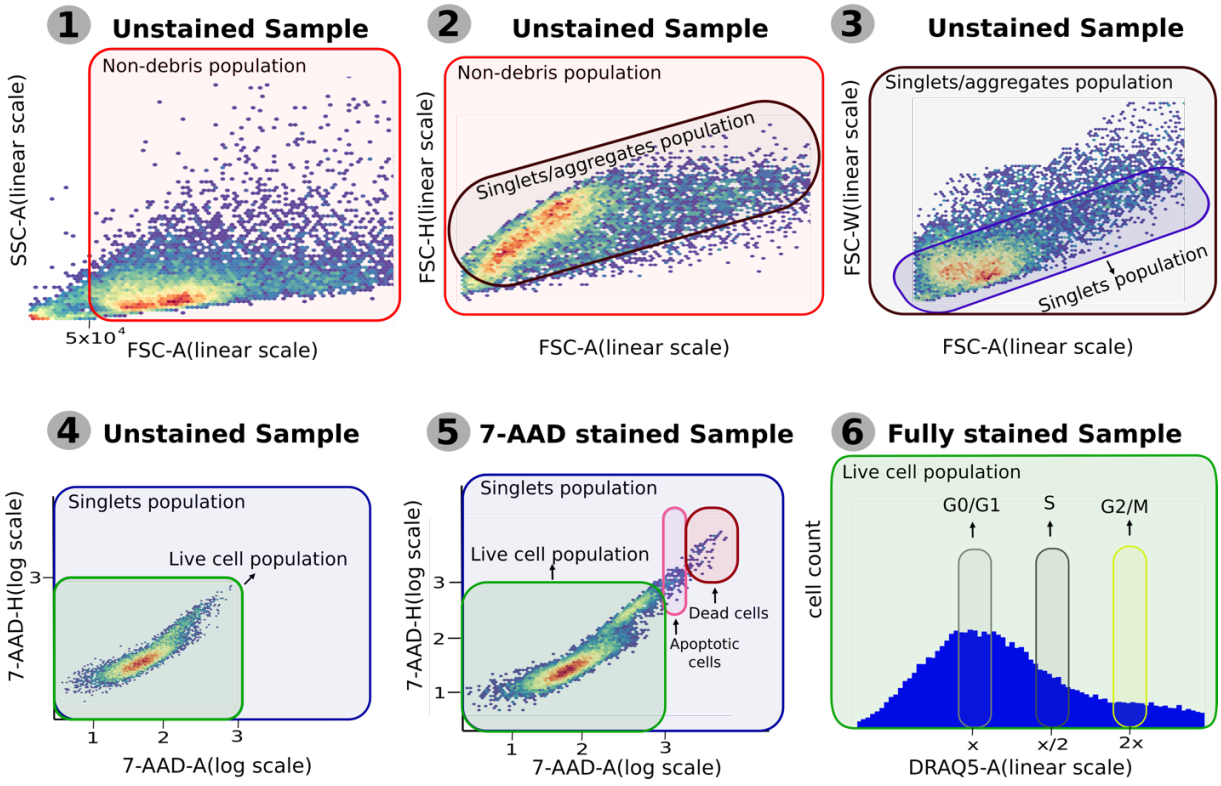

Figure 3.5: Schematic diagram illustrating the sequence of gates (filters) applied on the flow cytometry data in order to extract information only from single cells (1-3). This diagram also shows the gates for alive (G0/G1, S and G2/M subpopulations), apoptotic and dead cells (4-6). 1) Unstained cells are seen through the FSC-A vs SSC-A scatter plot to exclude debris. 2) The non-debris population is observed through the FSC-A vs FSC-H scatter plot to distinguish singlets from doublets. Cells with slightly less height are excluded as they are likely to be

doublets/aggregates. 3) The cells which were previously selected (a mixture of singlets and aggregates) are then examined using the FCS-A vs FSC-W plot to select the bottom half cells since singlets have a smaller width measurement. 4) The single cells selected from the unstained sample can then be plotted using the 7-AAD-A detection channel against the 7-AAD-H one. The area in which the cells appear is gated to select the 7AAD unstained cells (alive cells). 5) 7-AAD stained cells are then also observed through 7AAD-A and 7AAD-H channels to set up the dead cell and apoptotic gates. 6) Finally, by graphing a histogram of DRAQ5-A vs cell count for the alive cells, three gates are applied classifying the cells into G0/G1, S and G2/M populations. The width of these gates was selected to ensure a CV of the DRAQ5-A values less than or equal to 6% in all DNA populations (see section 1.4).

The sequence of gates which allow the selection of single cells (1 -3 in Figure 3.5) was applied in every experiment involving flow cytometric data. Once single cells were identified, these cells were then visualised through the detection channels of interest. Table 3.6 lists the relevant channels for each experiment conducted on the flow cytometer. These channels were used to collect data after the application of the gates for debris and aggregates exclusion.

Table 3.6: Detection channels used to extract data for statistical analysis for each of the experiments performed on the flow cytometer.

Experiment	Detection Channel
7AAD optimisation	7AAD-A, 7AA-H and 7AAD-W
DRAQ5 optimisation	DRAQ5-A, DRAQ5-H and DRAQ5-W
Sugar inhibition	LECTIN-A, LECTIN-H and LECTIN-W
Cell Culture Media Optimisation	SSC-A, SSC-H, SSC-W, FSC-A, FSC-H and FSC-W
Lectin Cytotoxicity	7AAD-A, 7AA-H, 7AAD-W, LECTIN-A, LECTIN-H and LECTIN-W
Variation of Cell Culture Parameters	7AAD-A, 7AA-H, 7AAD-W, LECTIN-A, LECTIN-H, LECTIN-W, DRAQ5-A, DRAQ5-H and DRAQ5-W

3.12.6 Experimental design

The purpose of a careful experimental design is to reduce sources of variability allowing the investigation of the effect caused by the desired inputs. The identification and management of the experimental factors involved in the cell culture process is an important step towards minimisation of unwanted variabilities. However, the standardization and calibration of the machinery used (see section 3.12.3) and sample preparation (see section 3.12.1) are also of paramount importance in the achievement of a high quality and reproducible biological data.

A cell culture process involves factors of 4 different natures: *inputs, controllable factors, uncontrollable factors and outputs* (Figure 3.6). Blocking and randomisation were adopted in the experimental work involving flow cytometric analysis. With the purpose of dealing with uncontrolled but observed inputs, blocking was employed, whereas randomisation was used to deal with uncontrolled and unobserved inputs. Controllable factors and relevant outputs were identified in order to investigate the input effects.

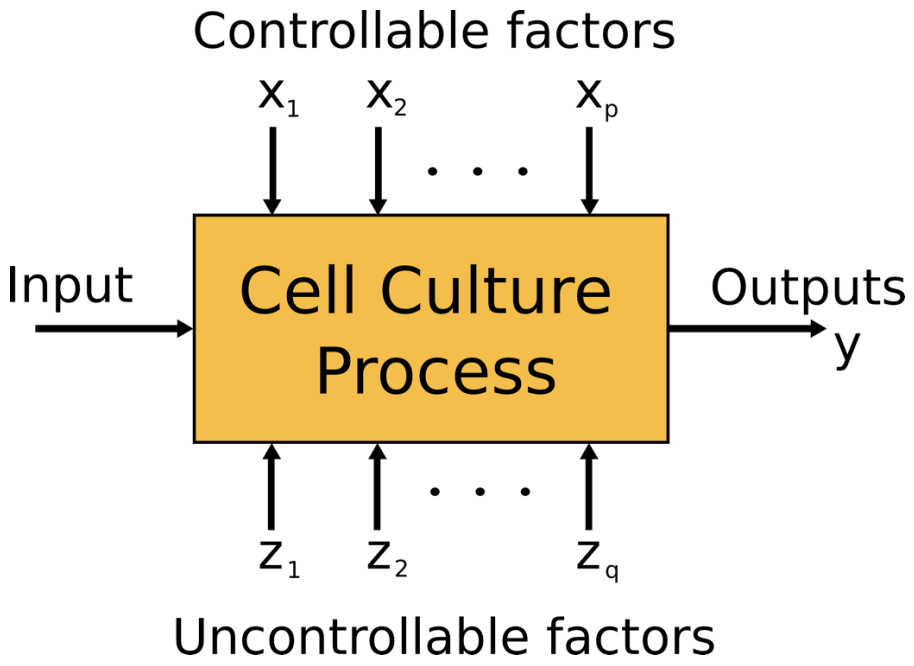

Figure 3.6: Four different types of factors involved in a general cell culture process: *Input, Controllable factors, Uncontrollable factors and Outputs.*

By blocking the number of passages of the cells in culture up to 10 passages after cell recovery (see section 3.11.6), any variation on the cell surface glycoprofile or cell metabolism which might be caused by the increase of the cell passage number was minimised. The use of DNA/cell viability dyes, 7-AAD (viability/DNA dye) and DRAQ5 (DNA dye), also blocked the cell population into groups so that the relevant outputs could be analysed per each group. These dyes allowed the decrease of data variability caused by the viability factor and DNA cell cycle factor within groups. Statistically, this blocking strategy allowed the comparison of the same groups from the different treatments with the baseline cells. For instance, the G0/G1 of the live baseline cells were then statistically compared to the G0/G1 of the live 32 °C cells of the temperature variation experiment (Figure 3.7).

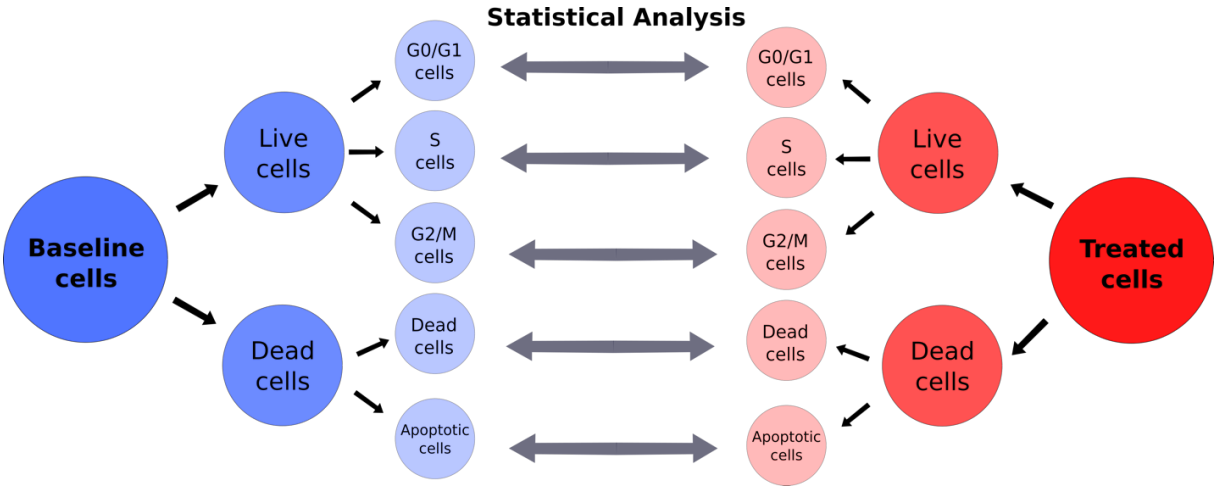

Figure 3.7: Diagram illustrating the statistical comparison between groups from treated and baseline cells. Such arrangement was allowed due to blocking cells into Live and Dead groups first, then live DNA cell cycle subpopulations (G0/G1, S, and G2/M), and fully dead and apoptotic cells.

Randomisation run was applied during the preparation of the samples for flow cytometry screening in order to deal with uncontrolled and unobserved variables. At the sample preparation step (see section 3.12.1) the microcentrifuge tubes were assigned a number. A numerical vector containing the ascending sequence covering the number of tubes was

created in R, then the *sample( )* command was run for this vector to obtain a random sequence. The lectins and DNA dyes were then added to the tubes following the random sequence generated. This strategy ensures that any variability due to the difference in the time of the application of lectins and dyes is spread across all the samples. The same sequence was also used at the data collection step.

While randomisation and blocking were used to deal with uncontrollable factors, controllable factors were identified to further minimise variabilities in the cell culture process. Temperature, CO_2, nutrient levels throughout the cell culture process, base medium and supplementation process, agitation speed, starting cell density, and cell sampling (the point in time in which cells are used for flow cytometric screening) are examples of factors which could be controlled and kept unchanged for baseline measurements. However, temperature, CO_2 and spent medium levels acted as inputs to obtain the measurements of the relevant outputs.

Temperature, CO_2, and nutrient levels throughout the cell culture process are factors which acted as input variables in a univariate study. In other words, each of these factors was varied at a time on the third day of the cell culture process. One day after the alteration, cells were then screened to measure the outputs.

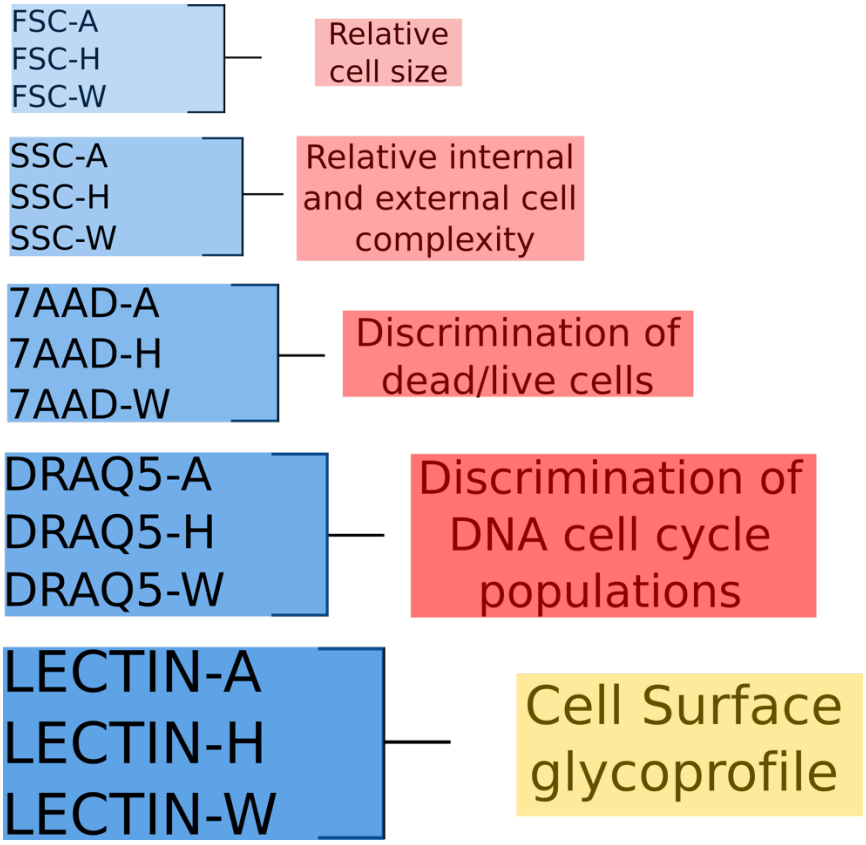

Figure 3.8: Schematic diagram summarizing the flow cytometry detector channels used as outputs and the information provided by each channel.

Several outputs were measured to investigate the effects of the inputs on the cell surface glycoprofile. With the aid of the flow cytometer, 15 outputs were measured which in turn provided information on different facets of the scientific work. For example, 7-AAD-A and 7-AAD-H provided information required to distinguish dead cells from the live ones, while DRAQ5-A was used to discriminate the cell population into the three main DNA cell cycle population: G0/G1, S, and G2/M. However, the most relevant information to address the main research question of this present work was obtained from LECTIN-A, LECTIN-H and LECTIN-W as lectin interactions on cell surface were measured through these detection channels (Figure 3.8). Measurements of the pH were also obtained as an output; however, this factor was measured with the aid of a pH electrode suitable for small quantities of cell suspensions (see section 3.12.1.8).

3.12.7 Statistical analysis

The roadmap for the statistical analysis was established based on the main research question of this scientific work: *Does the glycoprofile on CHO cell surface change as the cell is cultivated under stressful conditions?* From this question, the baseline of cell culture conditions was selected (37 °C, 5% of CO_2 and non-intervention of the nutrient level in the cell culture process) and different levels of temperature, CO_2 and spent medium were selected as stressful conditions. Only one factor was changed at a time, allowing the investigation of the influence of the factor on cell surface glycoprofile alone. For each variation of a factor, the response of the treatment on cell surface glycoprofile was statistically compared with the baseline cell surface glycoprofile.

When it comes to the assessment of the relationship between the baseline and a treatment, two experimental designs can be used: Longitudinal and Cross-sectional study design. Longitudinal requires that each data point (each screened cell) of the baseline sample is matched and related to a unique data point of treatment sample. These samples are then called paired-samples. However, in a cross-sectional study design, the data points in one sample (the baseline sample) are unrelated to the data points in the second sample (a treatment sample). These samples are then called independent samples (Rosner, 2000).

The cross-sectional study design was then selected as the most suitable one based on several aspects of the experimental work. On the other hand, the longitudinal study design was shown to be highly unsuitable based on the same aspects. In order to prepare cells for flow cytometric screening, multiple centrifuge and staining incubation steps are necessary (see section 3.12.1). Cells are exposed to a non-sterile environment during sample preparation and data collection. Therefore, due to these factors alone, the same cells are not suitable for re-culturing for later rescreening. Furthermore, it is not feasible to identify each individual cell to match the data from multiple treatments applied to it and the sample preparation steps are themselves stressful factors. For instance, lectins attached to the cells and DNA dyes which are incorporated by the cells during the sample preparation could have accumulative effects

on the glycoprofile of the cells, even if these cells had the required conditions to be re-cultured.

For the statistical treatment of the data collected through flow cytometry, the normal distribution model was used based on the central-limit theorem. This theorem states that for a large sample size ($n > 100$), the distribution of the observations can be assumed to be normal, even if the underlying distribution of individual observations in the population under investigation is not normal (Rosner, 2000). The arithmetic mean was used as the point of estimation of the centrality of the data distribution and the standard deviation as the estimator of the spread.

Therefore, the two-sample *t* test for independent samples with a significance level of 0.05 (α) was used for hypothesis testing (inferential analysis) and the computation of the *p*-values (Table 3.7). Since there was no reason to assume the equality of the underlying variances of the baseline and a treatment dataset, the *f* test was adopted first. Then the suitable *t* test was employed and power analysis was performed. While the *t* test calculated the levels of the statistical significance of the difference between the two samples (baseline and treatment), power analysis evaluated the likelihood of finding a significant difference when there was one (Figure 3.9).

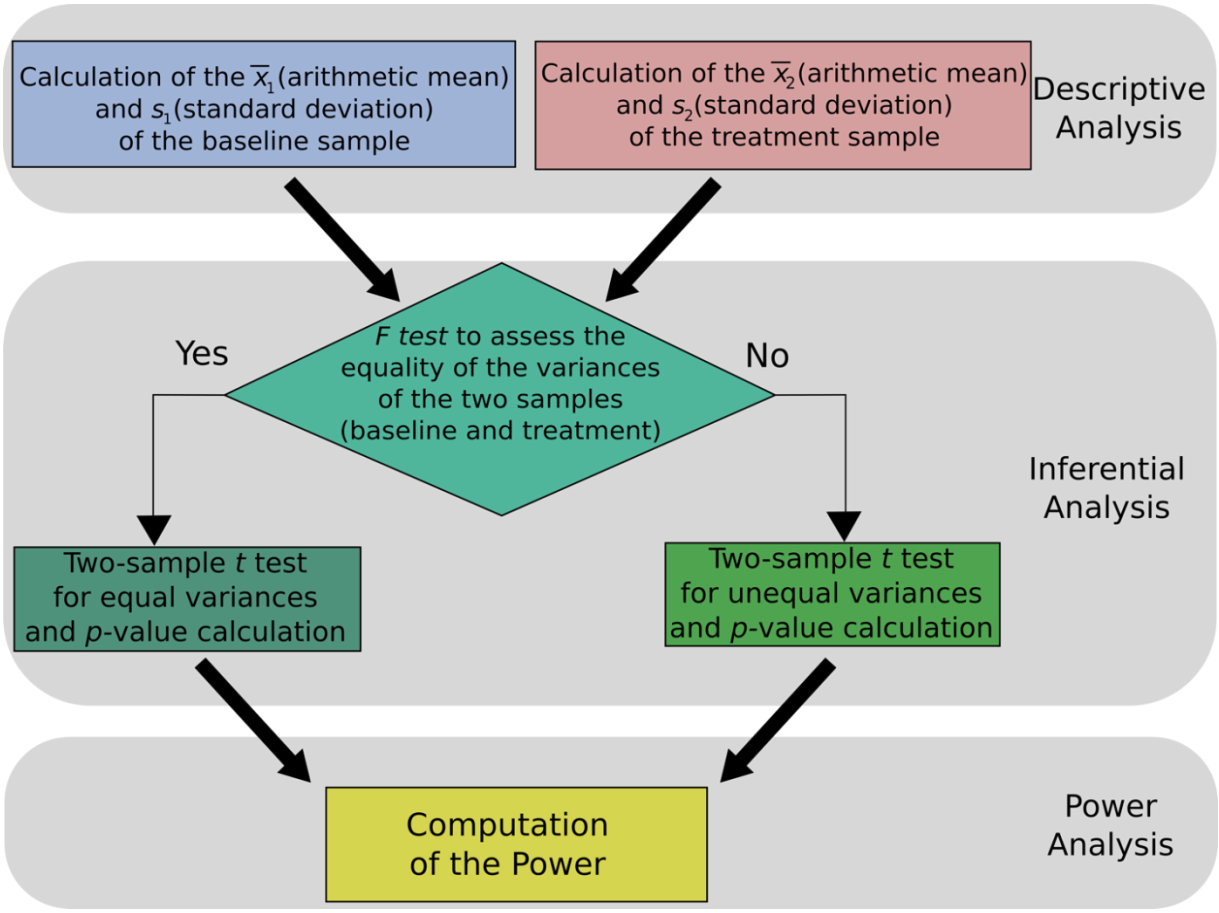

Figure 3.9: Statistical roadmap adopted for the analysis of flow cytometry data from the experiments involving the variation of the temperature, CO_2 and spent medium levels.

Table 3.7: *p*-values interval and the respective levels of statistical significance (Rosner, 2000).

***p*-value**	**Level of statistical significance**
$0.01 \leq p < 0.05$	*significant*
$0.001 \leq p < 0.01$	*highly significant*
$p < 0.001$	*very highly significant*
$p > 0.5$	*not statistically significant*
$0.05 \leq p < 0.10$	*there is a trend towards statistical significance*

3.12.7.1 The *F* Test

The significance test for the equality of two variances involves testing the hypothesis H_0: $\sigma_1^2 = \sigma_2^2$ versus H_1: $\sigma_1^2 \neq \sigma_2^2$. H_0, the null hypothesis, states that the two variances are equal whereas, H_1, the alternative hypothesis, states the variances are unequal. σ_1^2 and σ_2^2 are the true underlying variances of the two samples (baseline and treatment in the context of this present work). It was stated that the test statistic was based on the variance ratio of the samples s_1^2/s_2^2, where s_1 and s_2 are the standard deviations of the samples, followed an *F* distribution under H_0 with $n_1 - 1$ and $n_2 - 1$ *df* (degrees of freedom), where n_1 and n_2 are the sample sizes of the two samples. The *F* test was performed as a two-sided test, as it was intended to reject H_0 for both small and large values of s_1^2/s_2^2 (Rosner, 2000). The value 0.05 was the significance level (α) adopted for the *F* test which can be made more specific, as the Figure 3.10 illustrates.

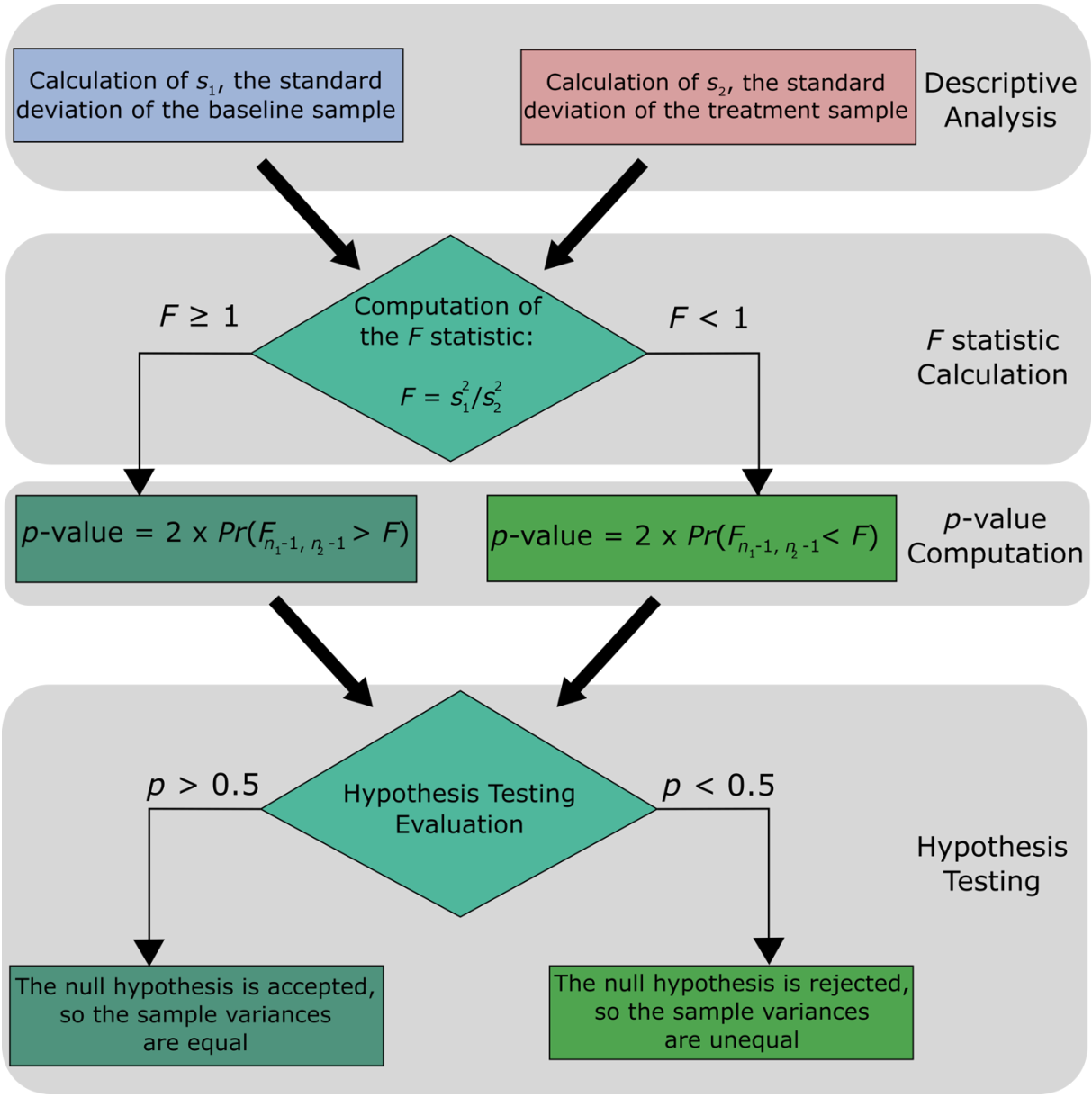

Figure 3.10: Statistical flowchart demonstrating the two-sided significance test for the equality of two variances. Where *Pr( )* means probability and *p* is the p-value (Rosner, 2000).

3.12.7.2 The independent two-sample *t* Test

The independent two-sample *t* test for the comparison of means involves testing the hypothesis H_0: $\mu_1 = \mu_2$ versus H_1: $\mu_1 \neq \mu_2$. H_0, the null hypothesis, states that the two means are equal whereas, H_1, the alternative hypothesis, states the means are unequal. The test was

conducted with 0.05 of significance level, t statistics and degrees of freedom were calculated according to the *F* test outcome (Figure 3.11). For samples with unequal variances, the Satterthwaite's method was used.

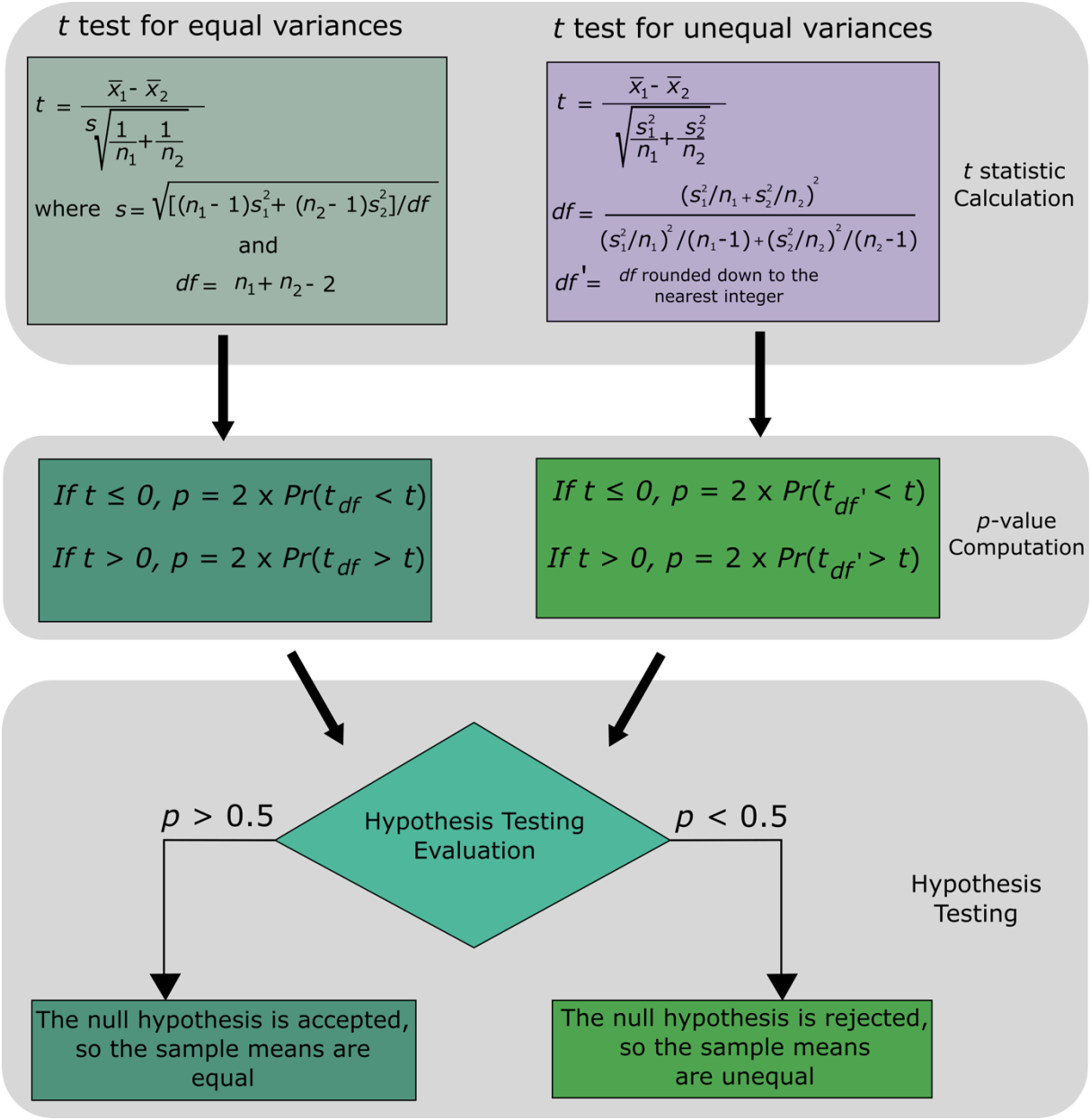

Figure 3.11: Flowchart of the two-sample *t* test to compare the means of the baseline and a treatment sample. Where $\bar{x}_1$ is the baseline mean, $\bar{x}_2$ is the treatment mean, n_1 is the baseline sample size, n_2 is the treatment sample size, s_1 is the baseline standard deviation, s_2 is the treatment standard deviation, *df* is the degrees of freedom, *df'* is the degrees of freedom ounded to the nearest integer, *p* is the *p*-value, and *Pr( )* is the probability (Rosner, 2000).

3.12.7.3 Power analysis

The power of a test provides information on the likelihood of detecting a significant difference provided that the alternative hypothesis is true; that is, provided that the mean of the baseline sample is different from the mean of a treatment sample. If the power is too low, there is little chance of finding a significant difference between the means of the samples and non-significant results are likely even if real differences exist. The power analysis was performed with 0.05 of significance level (α) (Equation 7) (Rosner, 2000).

$$Power = \Phi\left(-z_{1-\alpha/2} + \frac{\sqrt{n_1}|\bar{x}_1 - \bar{x}_2|}{\sqrt{s_1^2 + s_2^2/(\frac{n_2}{n_1})}}\right) \quad \text{(Equation 7)}$$

Where:

n_1 is the sample size of the baseline sample;

n_2 is the sample size of the treatment sample;

s_1 is the standard deviation of the baseline sample

s_2 is the standard deviation of the treatment sample

$\bar{x}_1$ is the arithmetic mean of the baseline sample

$\bar{x}_2$ is the arithmetic mean of the treatment sample

α is the test significance level set at 0.05

$z_{1-\alpha/2}$ is the statistic for inverse normal function at 1 – α/2

3.12.7.4 The statistical analysis of the lectin cell surface interaction across the DNA cell cycle

In order to evaluate the relationship between the DNA cell cycle with the cell surface lectin interaction excluding the cell physical dimension as an influential factor, the ratio of a lectin interaction parameter to the relative cell size parameter was calculated. In other words, the lectin interaction could then be evaluated across the subpopulations as a density variable. Since the ratio was obtained for each cell, the corresponding arithmetic mean and standard deviation were computed.

The descriptive statistical analysis of the ratios was performed across the DNA subpopulations (Go/G1, S, and G2/M) for all samples involving the variation of a cell culture parameter (see sections 3.12.2.2 and 3.12.2.3). Therefore, the statistical analysis allowed the investigation of the variation of the lectin interaction density on the surface of the cells as they go through the three distinct DNA cell cycle stages.

As the FSC-A provides information on the relative cell size and signal intensity, this channel was used as the parameter in the analysis of the ratios (see section 1.3). Likewise, LECTIN-A was selected allowing the determination of the ratio for each cell (LECTIN-A/FSC-A), the lectin interaction density, and the arithmetic mean of the ratios of a DNA subpopulation and the corresponding standard deviation.

3.12.8 Data Processing: R programming

The datasets obtained from the flow cytometric experiments were processed with the aid of R programming language. Bioconductor packages such as *flowWorkspace* and *flowCore* were used to read in and pre-treat the datasets on the R environment (RStudio Version 1.1.463). Algorithms to select the cells of interest, to manipulate the data and to perform the required calculations for the statistical analysis were then created (see the Appendix). For data visualization, *ggplot2* was the package adopted.

4 Results and Discussion

4.1 Lectin production and purification

AAL-2, LEC A and LEC B are recombinant lectins expressed with a histidine-tag (His-tag) to facilitate the purification process using an IMAC resin. These three proteins were developed by the research team and were included in the panel of lectins studied since they have shown relevant cell binding on CHO-DP12 cells in previous studies, and to compare the binding performance of these lectins with the commercial ones. The panel was also composed of four commercial lectins from Vector Laboratories: AAL (B-1395), WGA (B-1025), PNA (B-1075), and MAL II (B-1265).

AAL-2, LEC A and LEC B were produced in house as described in section 3.3. The purification was then conducted according to section 3.4.3 and SDS-PAGE was performed (see section 3.6) for the identification of purification fractions which contained the His-tagged protein isolated. Figure 4.1, 4.2 and 4.3 show, respectively, the gels with the purification fractions of AAL-2, LEC A, and LEC B.

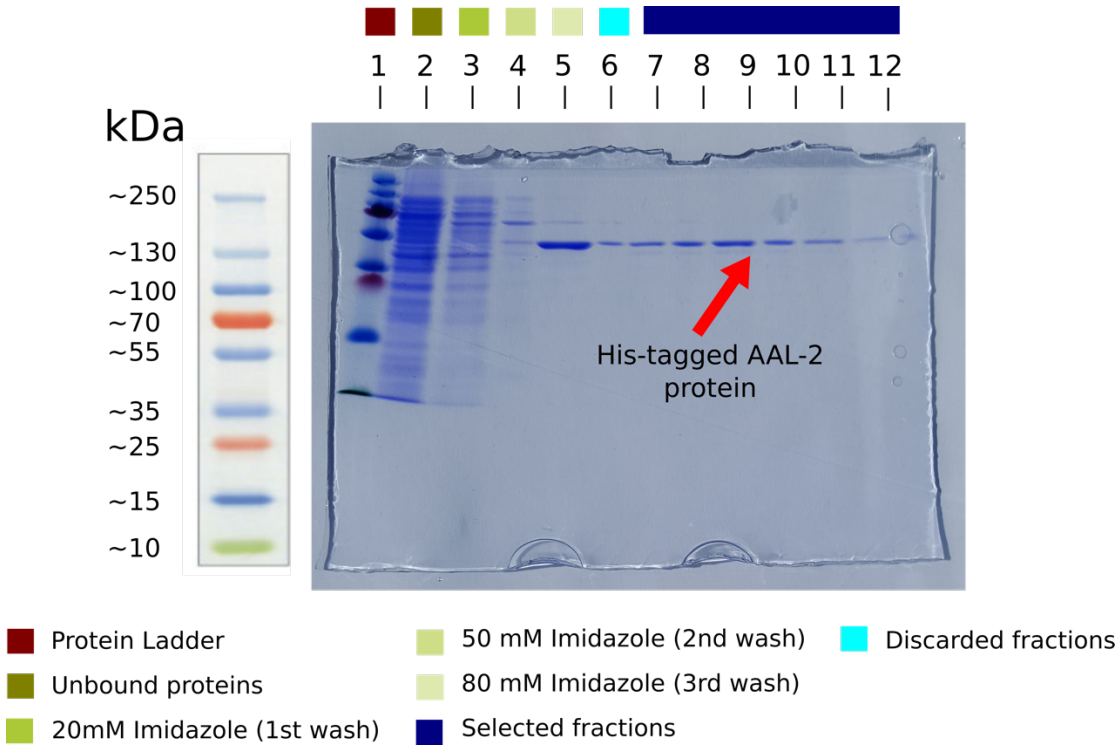

Figure 4.1: Purification of AAL-2. The figure shows a 15% SDS-PAGE gel on which samples taken at various stages of the process were fractionated. Lane 1: the PageRuler Plus Prestained Protein Ladder. Lane 2: unbound protein fraction. Lane 3: 20 mM Imidazole (first wash). Lane 4: 50 mM Imidazole (second wash). Lane 5: 80 mM Imidazole (third wash). Lanes 6-12: purification fractions (elution of the his-tagged protein using a 250 mM Imidazole buffer).

As shown in Figure 4.1, the his-tagged AAL-2 protein was effectively isolated after the application of Imidazole buffer solutions at three different concentration levels during the IMAC purification run. However, the first purification fraction (lane 5) was discarded, while the remaining ones (lanes 6-12) were selected as they contained a single band, the isolated His-tagged protein. The gel also indicates the approximate AAL-2 molecular weight which was between 35 and 55kDa bands of the protein standard ladder. Thus, the average 45kDa was estimated as the approximate AAL-2 molecular weight.

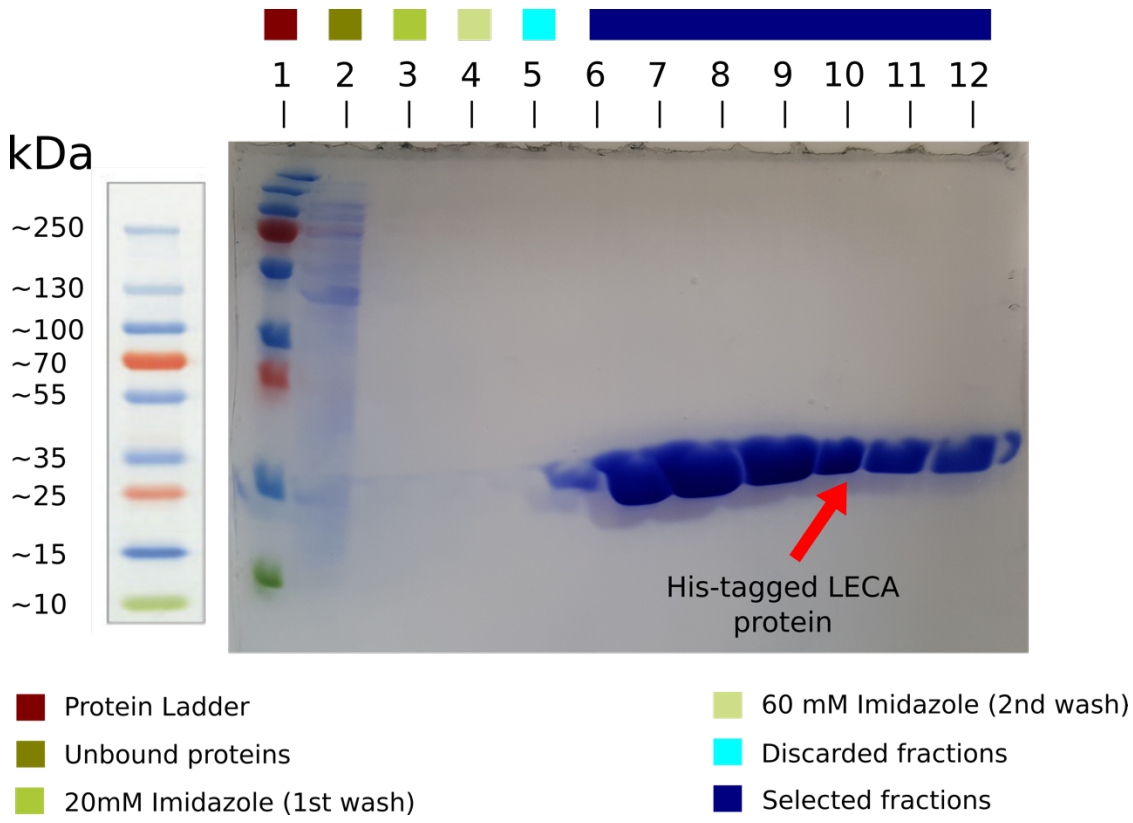

Figure 4.2: Purification of LEC A. The figure shows a 15% SDS-PAGE gel on which samples taken at various stages of the process were fractionated. Lane 1: the PageRuler Plus Prestained Protein Ladder. Lane 2: unbound protein fraction. Lane 3: 20 mM Imidazole (first wash). Lane 4: 60 mM Imidazole (second wash). Lanes 5-12: purification fractions (elution of the his-tagged protein using a 250 mM Imidazole buffer).

Figure 4.2 shows the effective purification of the His-tagged LEC A protein after the application of Imidazole buffer solutions at two different concentration levels during the IMAC purification run. Lanes 6 to 9 have revealed a thick His-tagged protein band indicating a high concentration level of the protein in these first purification fractions. Although a single band was observed in the first purification fraction (lane 5), the fraction was discarded, while the remaining ones (lanes 6-12) were selected as they contained the His-tagged LEC A protein in sufficient amount. The determination of the approximate LEC A molecular weight was directly obtained from the gel since the LEC A band was at the same level as the 15kDa band of the protein standard ladder.

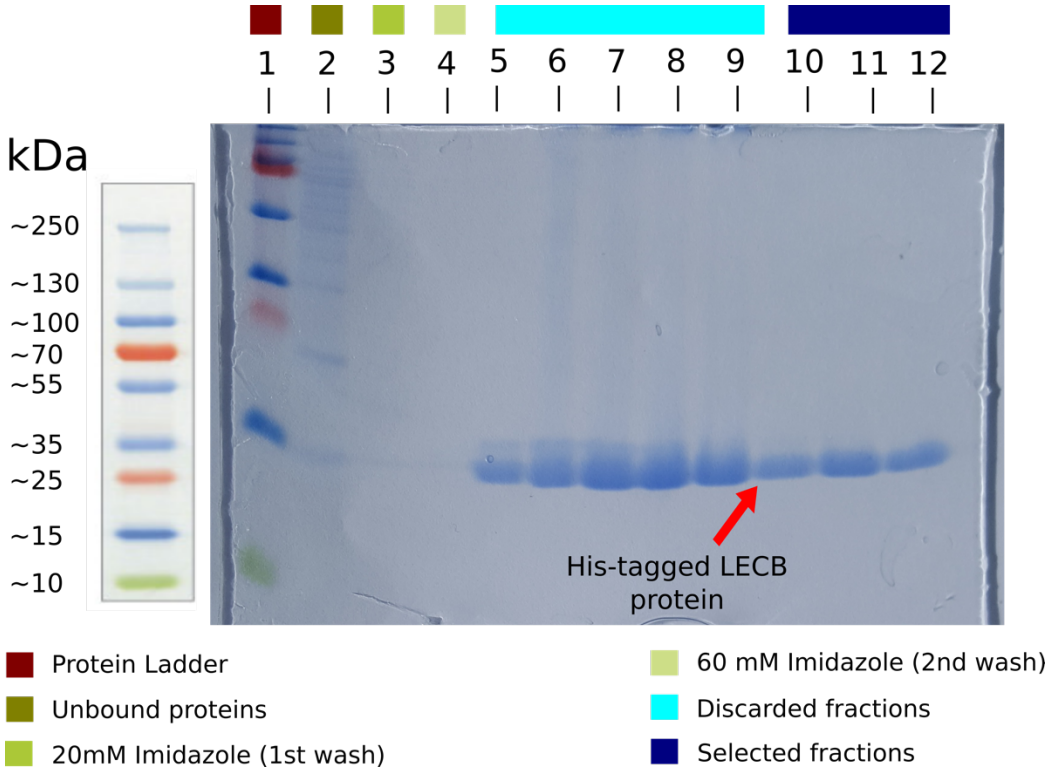

Figure 4.3: Purification of LEC B. The figure shows a 15% SDS-PAGE gel on which samples taken at various stages of the process were fractionated. Lane 1: the PageRuler Plus Prestained Protein Ladder. Lane 2: unbound protein fraction. Lane 3: 20 mM Imidazole (first wash). Lane 4: 60 mM Imidazole (second wash). Lanes 5-12: purification fractions (elution of the His-tagged protein using a 250 mM Imidazole buffer).

His-tagged LEC B protein was successfully isolated by the IMAC purification (see Figure 4.3). In addition, the determination of the approximate molecular weight of LEC B was obtained since the protein band was slightly below the 15kDa band of the protein standard ladder. Therefore, 14KDa might be a reasonable approximation of LEC B molecular weight.

In summary, AAL-2, LEC A and LEC B proteins were successfully purified using the IMAC purification technique. Also, the SDS-PAGE allowed the identification of the purest purification fractions and the determination of the approximate molecular weights of those lectins. AAL-

2 was determined to be the largest molecule (45kDa) in contrast to LEC B, the smallest one (14kDa). Whereas the molecular weight of LEC A was slightest larger (15kDa) than LEC B.

4.2 BCA assay: determination of lectin concentration levels

The BCA assay was employed with the purpose of determining a protein's concentration level in a sample. The purification fractions selected after assessment by SDS-PAGE, as discussed in the previous section, were pooled for each lectin generating three samples. The BCA assay was conducted as described in section 3.5 to firstly determine the protein concentration levels of those samples prior to biotinylation of the lectins. Then, BCA assay was performed again to determine the concentration of biotinylated lectins prior their use in ELLA analysis.

Lectin biotinylation was necessary since the fluorochrome of choice, the V450, is conjugated with streptavidin molecules enabling the fluorochrome to establish a strong bond with biotin molecules. The streptavidin has high affinity for biotin molecule allowing high quality flow cytometric analysis. ELLA analysis was then used to determine the biological activity and glycan specificities of the biotinylated proteins.

The BCA assay consists of first generating a standard curve which provides the mathematical model correlating absorbance values at 570 nm with the protein concentration levels of standard samples. The absorbance values of the standard samples were obtained whenever the determination of the protein concentration level of an unknown sample was required. Therefore, the standard and unknown samples were subjected to the same measurement conditions. Figure 4.4 illustrates the mathematical model obtained through the absorbance readings of standard samples allowing the determination of the protein concentration level of unknown samples of which the absorbance readings at 570 nm were known.

The second degree polynomial model demonstrated to be highly suitable for modeling BCA standard samples since the R-squared was almost 1 or 100%. Table 4.1 summarizes the concentration levels of the lectins before and after a biotinylation process. It can be observed that the concentration levels increased as a result of the biotinylation process.

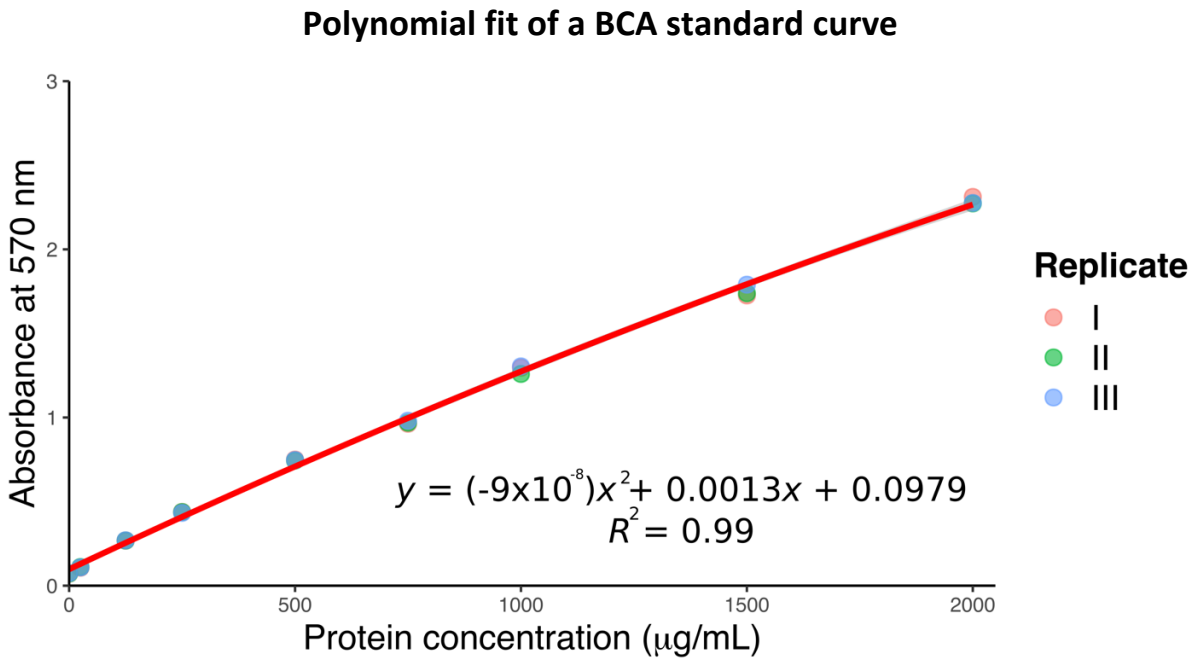

Figure 4.4: Second degree polynomial fit of a BCA standard curve. This model allows the determination of the protein concentration level of unknown samples of which the absorbance readings at 570 nm are known.

Table 4.1: Table summarising the concentration levels of non-biotinylated and biotinylated lectins determined from the BCA assay.

Lectin	Non-biotinylated (mg/mL)	Biotinylated (mg/mL)
AAL-2	0.28	1.1
LEC A	1.78	6.6
LEC B	3.11	5.9

4.3 ELLA: analysis of lectin biological activity and binding specificity

Following production and purification steps, it was necessary to assess the biological activity of the expressed proteins. Despite the fact these steps were closely monitored, they involve

multiple variables which can denature the proteins if these variables are not at the appropriate levels such as the level of temperature during production/fermentation, cell lysis and purification. Also, buffers which may have their pH altered for unforeseen reasons can cause damage to the proteins.

Furthermore, even the biotinylation process can also affect the lectin activity. Parameters such as the concentration of the Sulfo-NHS-SS-Biotin reagent could damage the proteins. Thus, ELLA analysis was performed to assess the biological activity of biotinylated lectins and provide information on their sugar binding specificities.

Figure 4.5 shows the determination of AAL-2 binding activity by ELLA. While AAL-2 has showed no considerable interactions with Fucose and Galactose since the absorbance (at 450 nm) values detected were very close to zero, it can be seen that the lectin demonstrated a strong interaction with N-Acetylglucosamine with an absorbance value around 0.18. Similarly, the commercial recombinant WGA lectin demonstrated a sugar specificity for N-Acetylglucosamine with an absorbance value around 0.25 confirming the information provided by the lectin supplier (see section 3.9).

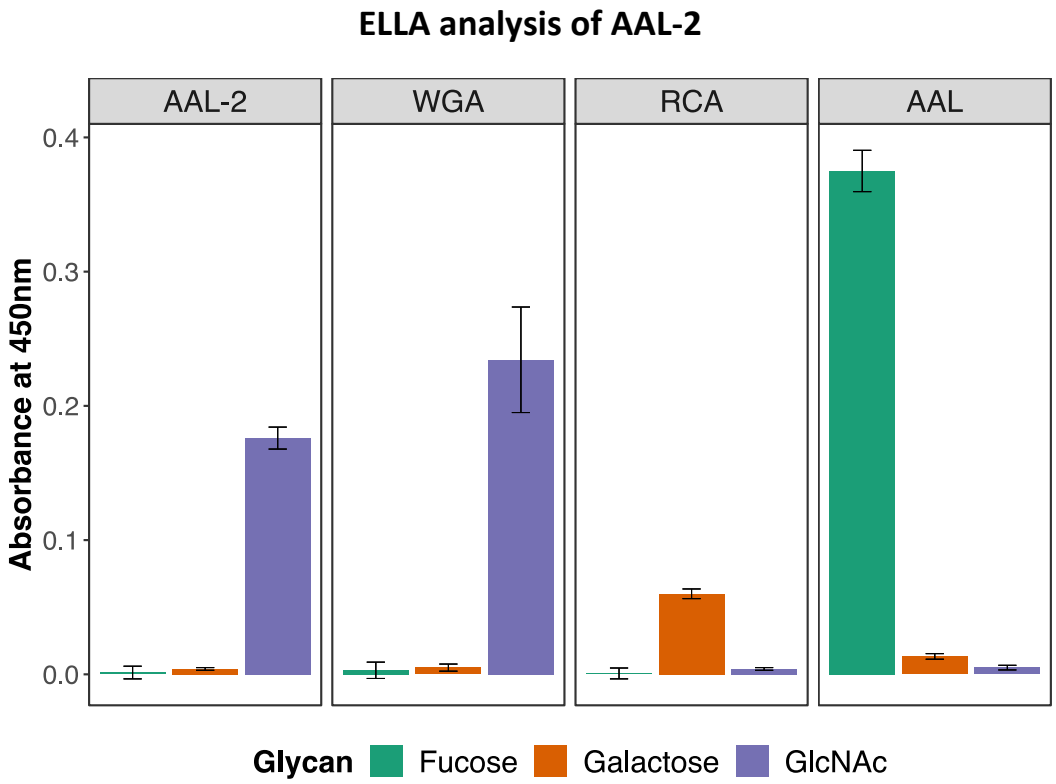

Figure 4.5: Determination of AAL-2 binding activity by ELLA. Bar chart which demonstrates the sugar specificity of AAL-2 and commercial lectin controls evidencing that these biotinylated lectins were biologically active. *Aleuria aurantia* lectin (AAL), *Aleuria aurantia* lectin 2 (AAL-2), Wheat germ agglutinin (WGA), and *Ricinus communis* agglutinin (RCA).

Also, the commercial lectins RCA and AAL which were used as negative controls for N-Acetylglucosamine specificity, confirmed its known sugar specificity; RCA for Galactose and AAL for Fucose, respectively. While AAL demonstrated a very strong interaction with its specific binding sugar as the absorbance value detected was around 0.36, RCA demonstrated a relatively weak interaction with Galactose.

The ELLA analysis of LEC B and LEC A can be observed through Figure 4.6 which shows absorbance values obtained from the interaction of these proteins with Fucose, Galactose and Mannose. Although LEC B and LEC A demonstrated considerable binding interactions with all sugars, LEC B showed a stronger interaction (higher absorbance values) than LEC A. However,

each of these lectins demonstrated a strong level of affinity for one of the sugars, LEC B for Mannose and LEC A for Galactose, respectively.

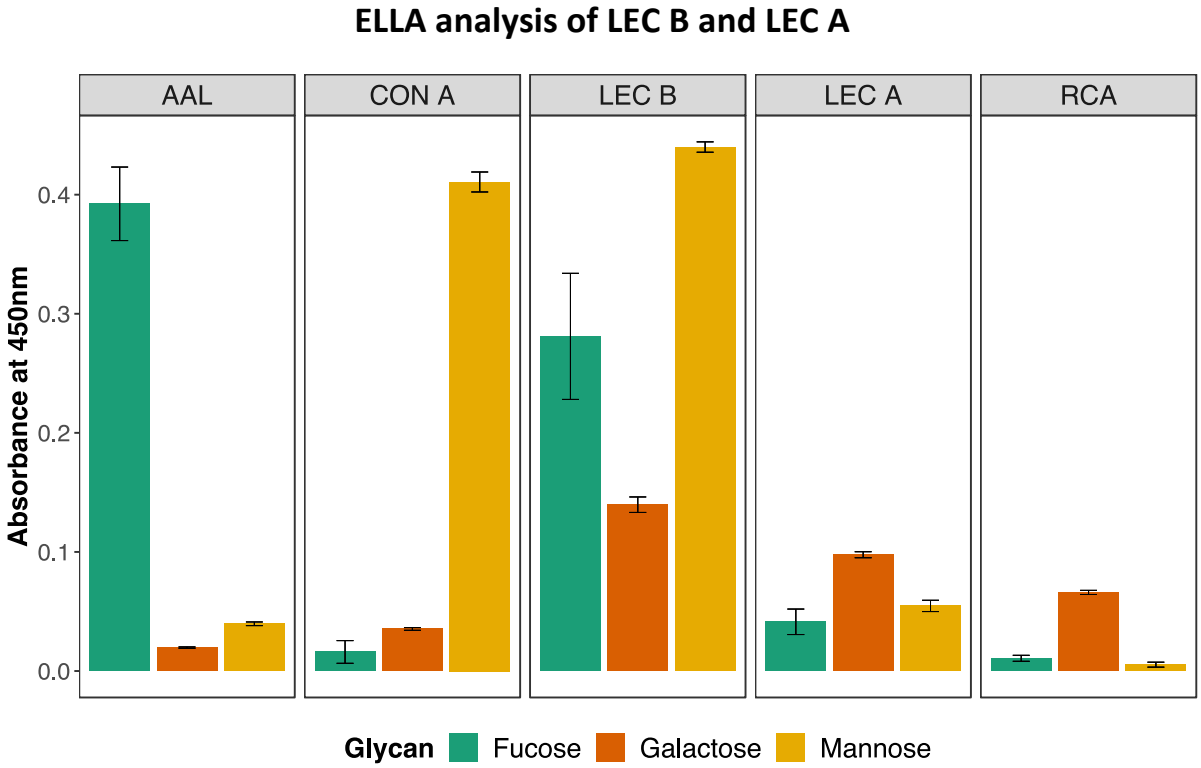

Figure 4.6: Lectin binding specificities. Bar chart demonstrating the sugar binding interactions of LEC B, LEC A and commercial lectin controls proving that these biotinylated lectins were biologically active. *Aleuria aurantia* lectin (AAL), Concanavalin A (CON A), and *Ricinus communis* agglutinin (RCA).

LEC A showed a similar sugar binding profile to RCA, whereas LEC B demonstrated a binding similarity to CON A. However, both RCA and CON A revealed a higher level of sugar binding specificity comparing to LEC A and LEC B.

In the case of CON A and LEC B, the absorbance value measured from the interaction between CON A and Mannose was around 0.42, while Fucose and Galactose absorbance values were below 0.05 for CON A. On the other hand, the interaction between LEC B and Mannose resulted in an absorbance value of 0.45, while for Fucose and Galactose the absorbance values

were around 0.27 and 0.15 respectively. Whereas in the case of LEC A and RCA, the latter demonstrated a high absorbance value for Galactose in relation to Fucose and Mannose, while LEC A did not. However, LEC B has showed a stronger interaction with Mannose than CON A. Similarly, LEC A showed a stronger interaction with Galactose than did RCA.

Since the lectins demonstrated specific interactions with sugar molecules, it can be concluded that recombinant lectins (AAL-2, LEC A and LEC B) as well as the commercial ones (WGA, RCA, AAL, and CON A) were biologically active proteins (Table 4.2).

Table 4.2: Table summarising the strongest sugar-binding molecule and the status of each lectin.

Lectin	Highest affinity sugar molecule	Status
AAL-2	N-Acetylglucosamine	Biologically active
WGA	N-Acetylglucosamine	Biologically active
LEC B	Mannose	Biologically active
CON A	Mannose	Biologically active
LEC A	Galactose	Biologically active
RCA	Galactose	Biologically active
AAL	Fucose	Biologically active

4.4 Experimental process optimisation: determination of process intervention and sample collection points

The experimental process optimisation study is part of a series of optimisation studies carried out prior the collection of data on CHO-K1 cell surface glycoprofile with the purpose of determining the optimal conditions allowing the minimization of data variability, therefore, increasing the data quality. Optimisation studies of the cell culture process and media were

conducted in order to determine the conditions for shortening the time required to carry out the experiments involving the variation of temperature, CO_2 and spent medium levels while allowing the reduction in data variability. As a consequence, the outcomes of these studies allowed the determination of the points of process intervention and sample collection for the aforementioned experiments. These points were selected within the stationary phase and while cell viability was above 90%. Thus, seeding density and the concentration of L-glutamine were varied to study the effect on the time needed for cell growth to achieve the stationary phase.

The cell growth curve is characterised by three main phases which initiates with the exponential phase when a cell population experiences growth at a rapid rate. Then, cell growth stabilizes allowing the cells to enter into the stationary phase. Finally, the death phase starts due to the exhaustion of nutrients and accumulation of toxins in the medium (Masters, 2000). Cell viability is a key cell culture parameter which measures the proportion of live cells in relation to dead ones. Thus, Figure 4.7 shows the relationship between the starting cell density, concentration level of L-glutamine and cell viability.

Figure 4.7 shows polynomial fit models for viability as a function of the cultivation time. Data variability is shown in the form of 95% confidence intervals. A 95% confidence interval is a range of values in which the likelihood of the true mean to fall in this range is 95%. Therefore, the width of the confidence interval illustrates the level of data variability. In other words, the wider the confidence interval, the higher the level of data variability.

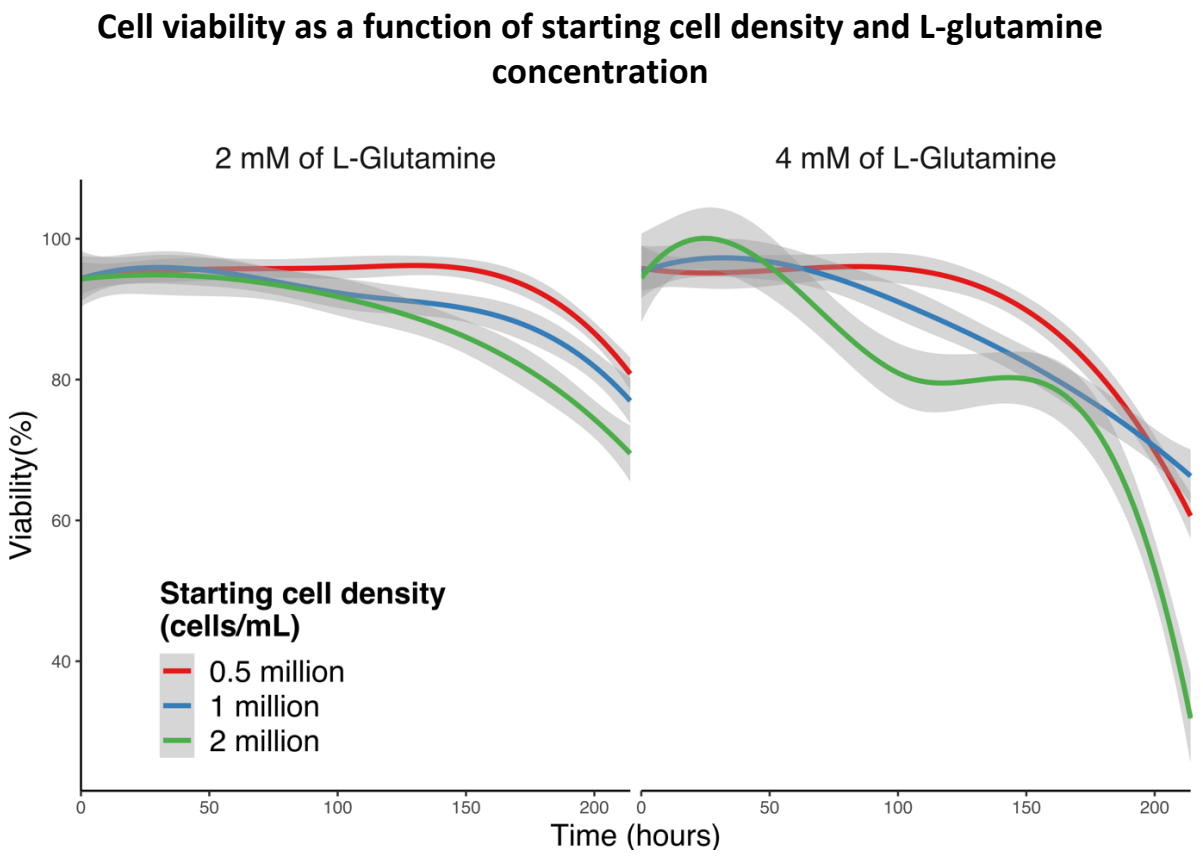

Figure 4.7: Line charts of polynomial fits with 95% confidence intervals of cell viability as a function of cell culture process time. The plot on the left is showing curves of cultures supplemented with 2 mM of L-glutamine and at 0.5, 1 and 2 million cells/mL of starting cell density. On the right, curves are shown for cultures supplemented with 4 mM of L-Glutamine and at 0.5, 1 and 2 million cells/mL of starting cell density. Data obtained through a BD FacsAria I flow cytometer using 7-AAD to stain dead cells (see section 3.12.1.1).

It can be observed through Figure 4.7 that the variability in the data obtained from 4 mM L-glutamine cultures was higher than when 2 mM L-glutamine was used. Furthermore, during the first half of the cell culture process, 4 mM L-glutamine curves decreased in viability at a higher rate. As the starting cell density increased, the rate at which cell viability dropped increased. Cultures with 4 mM L-glutamine seeded at 2 million cells/mL, in particular, showed a steep decrease in the viability up to 100 hours of the process. Cultures supplemented with 2 mM L-glutamine also revealed the same pattern. However, none of its cultures demonstrated a dramatic decrease in cell viability throughout the entire process.

Since the curve at 1 million cells/mL seeding density which was supplemented with 4 mM of L-glutamine and the curve at 2 million cells/mL with 2 mM of L-glutamine have demonstrated similar profiles, the conditions of these two curves could then potentially optimize the process of cell culture and media. To further evaluate the experimental data, pH measurements throughout the cell culture process time were taken.

As described in section 3.12.1.1, two sets of cell culture tubes were monitored in parallel. The first set was used for daily sampling for flow cytometric analysis and pH measurements, while the other was only used to collect samples for pH measurements. The sets were labelled as FC/pH and pH respectively (Figure 3.3).

Values of pH were obtained with the aid of a pH electrode suitable for cell suspensions. However, this electrode could have an effect on the cells influencing the viability outcomes. Therefore, the pH measurements of FC/pH samples were taken only after flow cytometric readings. Prior to flow cytometric analysis, 1 µL of 7-AAD and 50 µL of CountBright™ were added to those samples. Since the effect of these compounds on the pH of the samples was unknown, setting up a parallel set of cell culture tubes for pH monitoring only was necessary. Figure 4.8 shows the polynomial fits of the pH readings of the cultures seeded at 1 million cells/mL with 4 mM L-glutamine and at 2 million cells/mL with 2 mM L-glutamine.

It can observed through Figure 4.8 that the pH measurements of FC/pH samples ranged from around 7.5 to 8 pH and 7.3 to 8 pH for 2 mM of L-glutamine and 4 mM of L-glutamine cultures respectively. Whereas the ranges of pH samples were 7.1 to 8 pH and 6.9 to 8 pH for the 2 mM of L-glutamine and 4 mM of L-glutamine cultures, respectively. Therefore, the addition of 1 µL of 7-AAD and 50 µL of CountBright™ could have an influence on the pH of the samples by decreasing pH measurements. As a result, the pH readings of pH samples were used to further evaluate the experimental data.

pH monitoring of the *FC/pH* and *pH* samples across cell culture process time

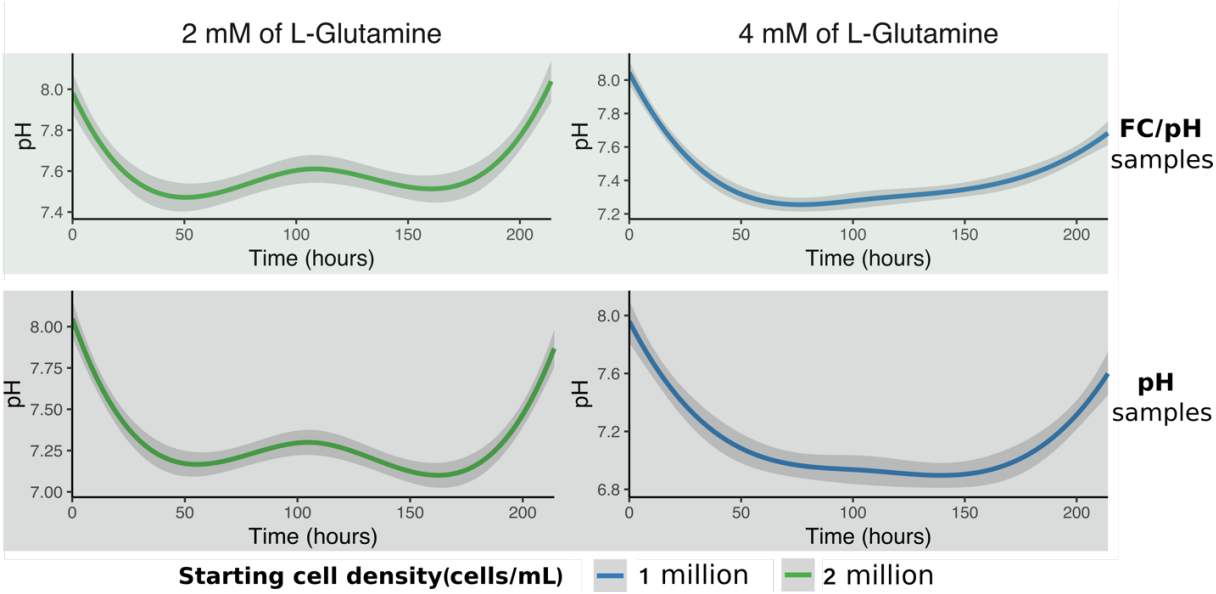

Figure 4.8: Line charts of polynomial fits with 95% confidence intervals of pH as a function of cell culture process time. The top plots show curves for **FC/pH** samples: the upper left plot shows the culture supplemented with 2 mM of L-glutamine at 2 million cells/mL starting cell density, while the upper right plot shows the culture supplemented with 4 mM of L-glutamine at the seeding density of 1 million cells/mL. The bottom plots show curves for **pH** samples: the lower left plot shows the culture supplemented with 2 mM of L-glutamine at 2 million cells/mL starting cell density, whereas the lower right one shows the culture supplemented with 4 mM of L-glutamine at the seeding density of 1 million cells/mL. The values of pH were obtained using an Orion Semi-micro pH electrode (see section 3.12.1.1).

The bottom plots of Figure 4.8 (pH samples) revealed that the variability, shown by 95% confidence interval, of both curves was quite similar. Nevertheless, data from the culture with 2 mM of L-glutamine and at 2 million cells/mL seeding density showed a substantial reduction in its 95% confidence interval width at the first and last 50 hours of the process. This pattern is also observed in the data of 4 mM of L-glutamine and 1 million cells/mL culture; however, its variability did not reduce at the same extent. Furthermore, the latter culture ranged from 6.9 to 8 in pH, while the former ranged from 7.1 to 8. Consequently, the culture with 2 mM of

L-glutamine and at 2 million cells/mL of seeding density showed the minimisation of pH variability throughout the cell culture process.

Although the shortening of the cell culture process time was the primary goal of this particular experiment, the minimisation of the variability of the cell culture process parameters was one of the main experimental design objectives in this research work. Therefore, the culture with the media supplemented with 2 mM of L-Glutamine and seeded at 2 million cells/mL was demonstrated to be the most suitable one.

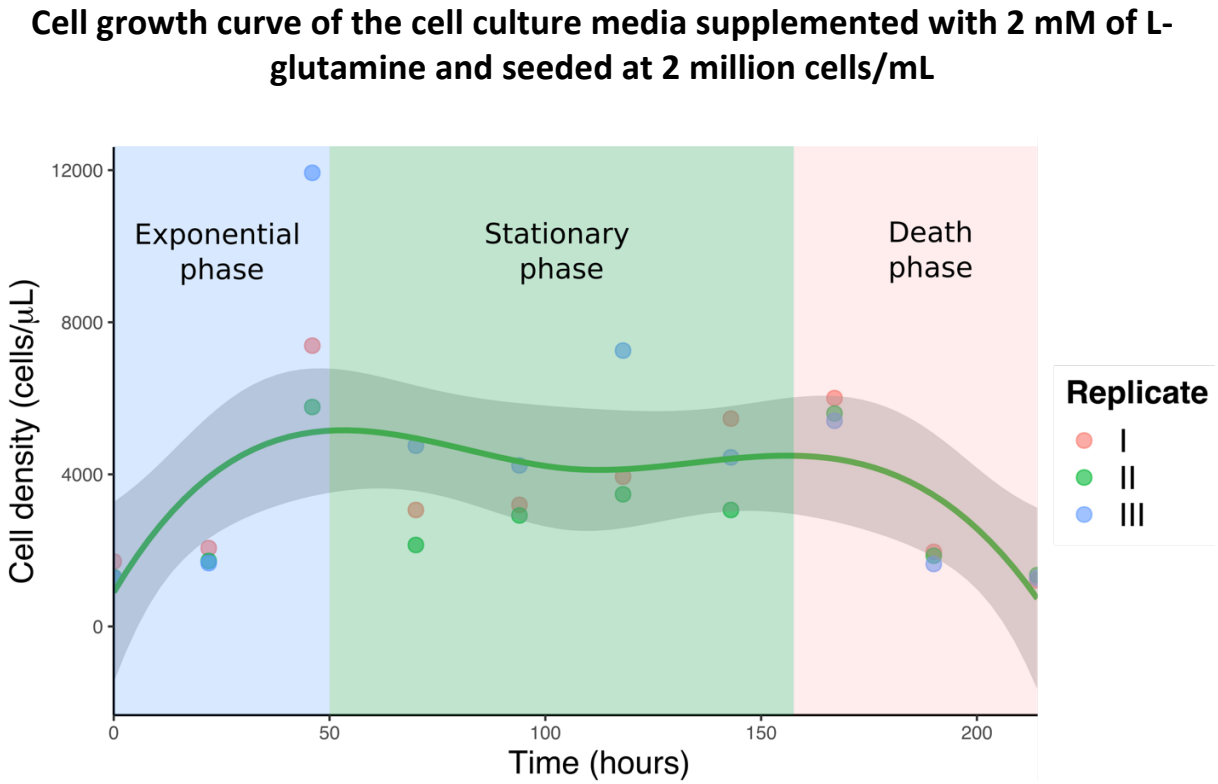

Figure 4.9: Line chart of a polynomial fit with 95% confidence interval of the cell density of viable cells of the culture supplemented with 2 mM of L-glutamine and seeded at 2 million cells/mL. The figure shows the three stages of the cell growth: the exponential or log phase, the stationary and the death one. Data obtained through a BD FACSAria™ I flow cytometer using CountBright™ absolute counting beads to determine cell density and 7-AAD to stain dead cells (see section 3.12.1.1).

The cell growth curve (viable cell density) of the optimal culture can be observed in Figure 4.9. As previously described in this section, the three main phases of the cell growth can be clearly distinguished in the figure. The first phase, the exponential or log one, takes place throughout the first 50 hours of the culture. Then the cell population enters into the stationary phase which lasts for over 100 hours. And finally, the death phase is initiated causing a rapid decrease in the density of the viable cell population.

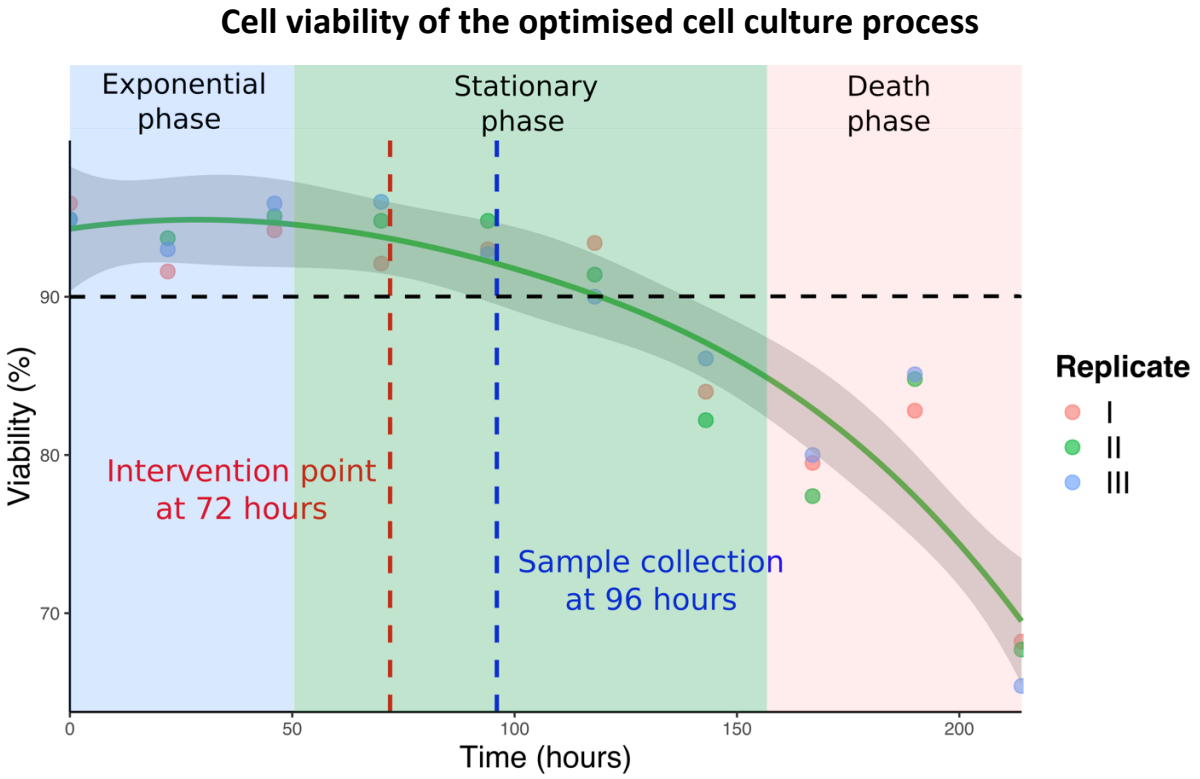

Figure 4.10: Line chart of a polynomial fit with 95% confidence interval of the viability of the culture supplemented with 2 mM of L-glutamine and seeded at 2 million cells/mL. The figure highlights the three stages of the cell growth: the exponential or log phase, the stationary and the death one. Additionally, the points of process intervention and sample collection can be observed as well as the 90% viability threshold (black dashed line). Data obtained through a BD FACSAria™ I flow cytometer with cells stained with 7-AAD to determine viability (see section 3.12.1.1).

After the identification of the stationary phase, the viability curve of the optimal culture was then closely observed in order to determine the points of process intervention and sample collection. Figure 4.10 shows the culture polynomial fit of cell viability as a function of process time.

The intervention point was selected at 72 hours from the seeding time and samples were collected for flow cytometric analysis 24 hours after, that is, at 96 hours from the seeding time. Although the confidence interval is partly below the black dashed line (90% viability) at the sample collection point, most of the interval is above the line. Therefore, the majority of the samples collected at the 96 hour point is highly likely to have cell viability above 90%.

4.5 DNA dyes optimisation

Two DNA dyes were used in the majority of the flow cytometric analyses to both discriminate live cells from dead ones and to split live cells into the three main DNA cell cycle subpopulations: Go/G1, S, and G2/M. While the 7-AAD fluorescent reagent was used to discriminate live cells from the dead ones, the DRAQ5 fluorochrome was employed to identify the DNA cell cycle of live cells.

With the purpose of determining the incubation time and concentration levels of the dyes, which allowed the detection of a relative strong fluorescence signal from these molecules while reducing the signal variability inserted into the overall data, different concentration levels of the dyes and different incubation periods of time were investigated. The ranges of variation of the incubation time and concentration levels were established based on the protocol for flow cytometric analysis suggested by the suppliers of these DNA dyes.

4.5.1 7-AAD

As suggested by the 7-AAD supplier (ThermoFisher Scientific, A1310), cells were first incubated for 15 minutes to evaluate the effect of different concentration levels on the strength of 7-AAD fluorescence signal and the variability of the data. Figure 4.11 shows a box plot superimposed with a violin plot. While the box plot summarises the shape of the distribution showing the lower (25^{th} percentile), median (50^{th} percentile) and upper quartiles (75^{th} percentile), the violin plot is a compact representation of the "density" of the distribution, highlighting the areas where more points are found. The figure demonstrates the data distribution and the median for five different concentration levels of 7-AAD. The concentration levels are being measured in microliter volumes of 7-AAD per 0.5 mL of cell suspension containing 2 million cells in fully supplemented pre-warmed medium, as described in sections 3.12.1.2 and 3.12.1.3.

Figure 4.11 reveals that the distributions of the data obtained for all 7-AAD concentrations are positively skewed since the upper quartiles are farther from the medians than the lower quartiles. Additionally, the median (a measure of central tendency) and the range between the quartiles (the spread or the variability of the data) increases as the concentration of 7-AAD increases up to 4 μL. However, the data obtained from the cells incubated with 5 μL shows a decrease both in the median and the range between the quartiles, showing that further increases in the 7-AAD concentration might reduce the fluorescence signal strength rather than increase it.

Although the median rises as the 7-AAD volume increases up to 4 μL, the change is not substantial; thus, the strength of the fluorescence signal detected from 7-AAD-A detector channel is not significantly affected by the increasing concentration of 7-AAD. In contrast, data variability, which can be observed through the shape of the violin plot and the distance between the lower and upper quartiles as well as the distance between the whiskers of the box plot (smallest nonoutlying and largest nonoutlying values), rises as the concentration of 7-AAD increases, except for the concentration level resulted by adding 5 μL volume of the dye in which data variability and median are slightly lower than when 4μL volume is added.

Consequently, 1 μL volume 7-AAD demonstrated to be the most suitable quantity to be added to a 0.5 mL of CHO-K1 cell suspension containing 2 million cells. This means 2μg/mL as final concentration of 7-AAD.

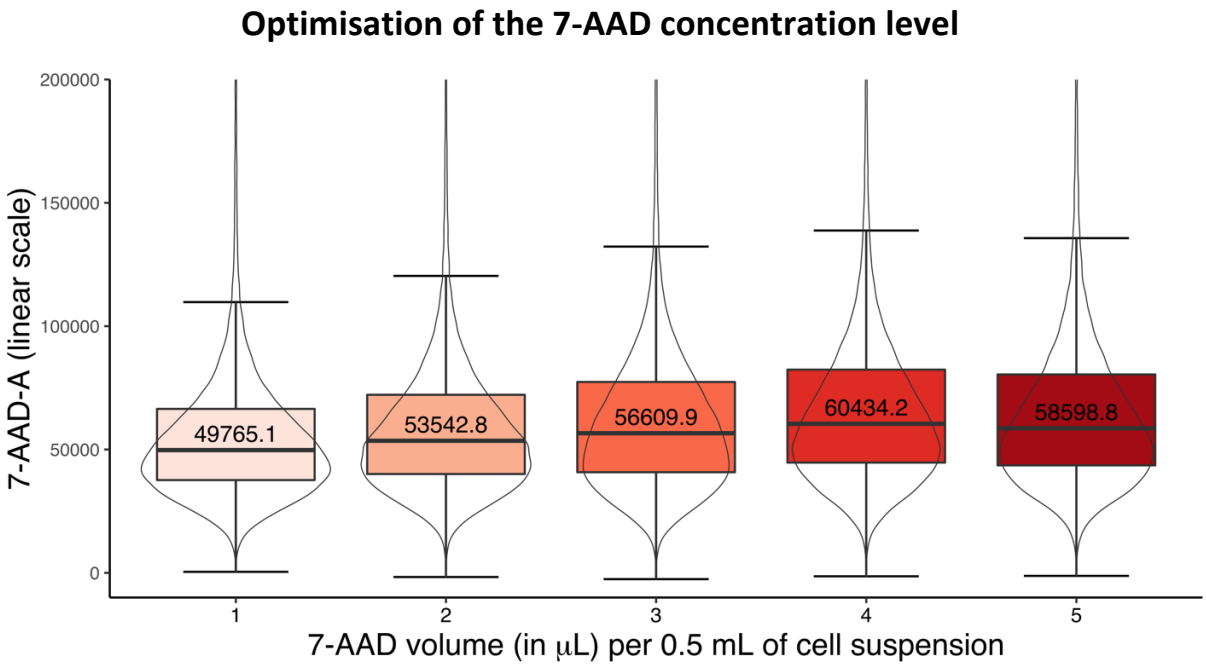

Figure 4.11: Box plot overlaid with a violin plot showing the data distribution and the median of the fluorescence signals of cells incubated for 15 minutes at different concentration levels of 7-AAD. The concentration is measured in terms of 7-AAD volume (μL) per 0.5 mL of cell suspension containing 2 million cells. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.2).

Following the determination of the most suitable concentration level of 7-AAD, data was obtained from three different periods of incubation time using this concentration level. Subsequently, data was also visualised through a box plot overlaid with a violin plot.

The visualisation of the data (Figure 4.12) demonstrates that the distributions are all positively skewed given the upper quartile is farther from the media than the lower quartile. This can be

observed in the box plot as well as in the shape of the violin plot, showing that the majority of the data points are closer to the lower quartile.

The plots also show that the fluorescence signal strength measured on the 7-AAD-A channel does not alter significantly as the cells are incubated with 7-AAD for longer periods of time, but signal strength slightly declines. This is clearly observed by evaluating the values of data distribution median: 68029.9, 65049.9 and 65562 regarding to 5, 10 and 15 minutes respectively.

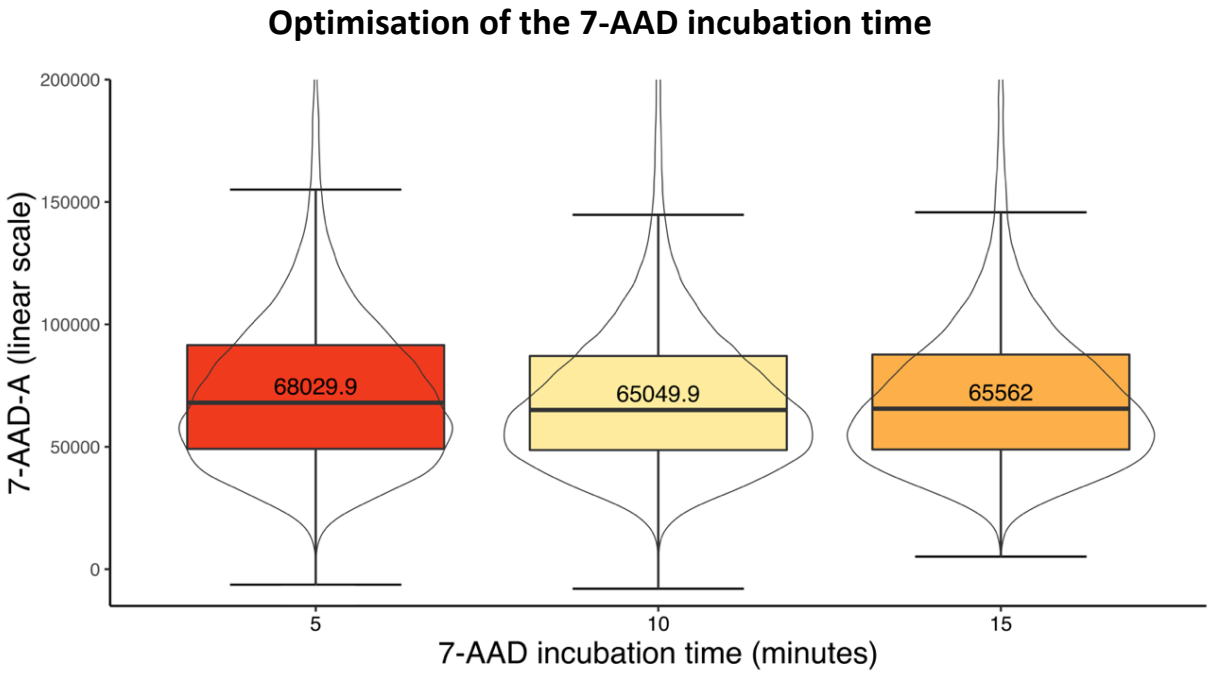

Figure 4.12: Box plot overlaid with a violin plot showing the data distribution and the median of the fluorescence signals of cells incubated with 1 µL of 7-AAD at different incubation times. The 7-AAD volume was added to 0.5 mL of cell suspension containing 2 million cells. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.3).

However, by incubating the cells for longer periods of time, the variability reduces as can be observed by the increasing definition of the data distribution peak illustrated by the shape of the violin plot, and the distances between the whiskers and between the upper and lower

quartiles highlighted by the box plot. Despite the fact the distance between the quartiles data obtained at 10 and 15 minutes seems to be equal, the distance between the whiskers of the 15 minute dataset and the shape of its violin plot clearly show the reduction in the spread of the data. As a result, the 15 minute incubation time is the most suitable since the fluorescence signal was sufficient for detection while providing the greatest reduction in the variability of the data.

Therefore, the 1 μL of 7-AAD volume and 15 minutes of incubation were the most suitable conditions to obtain high quality of data for 2 million CHO-K1 cells suspended in 0.5 mL of fully supplemented pre-warmed medium. In conclusion, 7-AAD optimisation experiments allowed the successful determination of the 7-AAD volume and its incubation time which provided sufficient detection of the fluorescence signal emitted by the dye while reducing the variability of the data.

4.5.2 DRAQ5

With the purpose of staining cells with DRAQ5 for DNA cell cycle analysis by flow cytometry, BioStatus, the supplier of the DRAQ5 (stock solution at 5mM, DR50200), recommends up to 10 μL of the dye for 200,000 cells resuspended in 0.5 mL of a buffer solution such as PBS. A range from 5 to 30 minutes of incubation time at room temperature was recommended, but the supplier suggests that the incubation time can be accelerated at 37°C. Since no recommendation is given for staining 2 million cells in 0.5 mL of cell culture media, DRAQ5 optimisation experiments initially investigated the effect of 1, 2, 3, 4 and 5 μL of DRAQ5 volumes per 0.5 mL of medium containing 2 million cells for 20 minutes incubation time at 37°C. Subsequently, the periods of incubation time of 5, 10, 15, 20 and 25 minutes at 37°C were tested for the optimal concentration of DRAQ5 using the same number of cells and medium volume (see sections 3.12.1.4 and 3.12.1.5).

Figure 4.13 shows two different plots with the data obtained from different volumes of DRAQ5. The top plot shows data distributions in the form of a boxplot, while the bottom one

shows the distribution density, highlighting the regions where most of the points are found. The data distributions are positively skewed for all DRAQ5 concentration levels. Data variability as well as the median distribution increases with DRAQ5 concentration. In other words, the increase in DRAQ5 concentration causes a rise in the strength of the fluorescence signal and in the spread of data.

Despite the fact that the DRAQ5 fluorescence signal is considerably strengthened with higher concentration levels of the dye, the gain in data variability compromises the quality of the data. The fluorescence signal for 1 μL demonstrated to be sufficient for detection with the advantage of providing the smallest level of data variability. Furthermore, the well-defined peak of the 1 μL dataset enhances the ability to detect the Go/G1 DNA cell cycle population which is located by the highest peak observed in a density plot of a DNA dye (see section 1.5).

The experiment to analyze the effects of different periods of DRAQ5 incubation time was then conducted by adding 1 μL of the dye (10μM a final concentration of DRAQ5). The datasets can be visualised in Figure 4.14. This figure is also composed of complementary plots, the box plot (top) and the violin plot (bottom), demonstrating that the distribution of all datasets are positively skewed; therefore, most of the data points are found below the median value. Also, as can be clearly seen in the top plot, the increasing incubation time diminished the strength of the fluorescence signal emitted from DRAQ5. Although the distribution median of the data obtained for 25 minutes (63873.9) is nearly half of the 5 minute distribution median (119842.6), the variability of the former is considerably lower than the latter. However, the signal detected from 25 minute variation proved to be strong enough for detection. As can be observed in the bottom plot, this incubation time demonstrated to be the most suitable one, given its sufficient signal strength and the greatest reduction in data variability. Furthermore, the reduction in data variability improves the ability to identify the highest peak of a DRAQ5 density plot, facilitating the location of the fluorescence signal from the Go/G1 cell population.

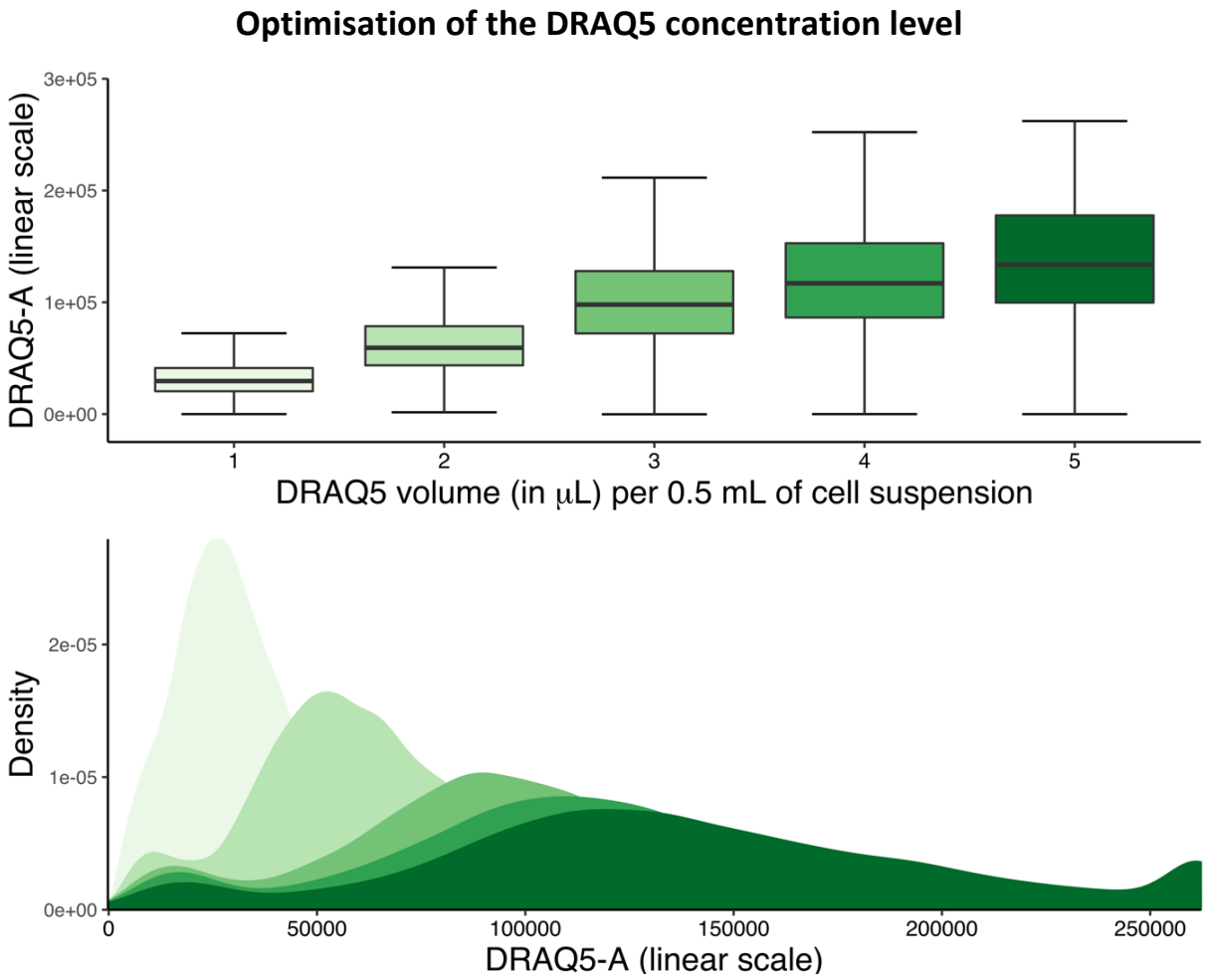

Figure 4.13: Box plot (top) and density plot (bottom) showing the distribution of the data obtained for cells stained with different volumes of DRAQ5 in 0.5 mL of medium containing 2 million cells. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.4).

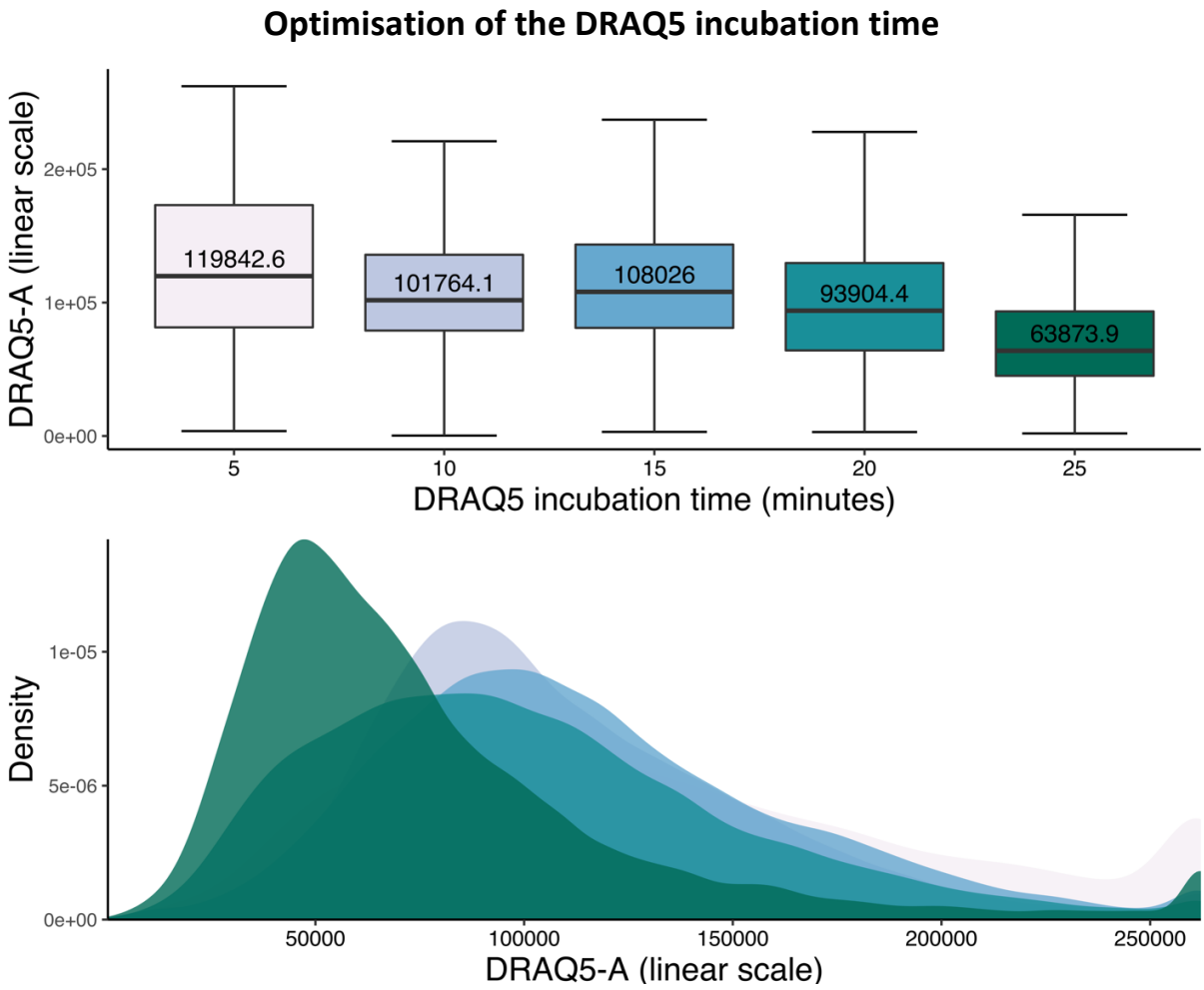

Figure 4.14: Box plot (top) and density plot (bottom) showing the data distributions of the datasets obtained for cells incubated for increasing periods of incubation time. All the cells were stained with 1 µL of DRAQ5 in 0.5 mL of media containing 2 million cells. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.5).

To summarise, the 1 µL of DRAQ5 volume and 25 minutes of incubation at 37°C were the most suitable conditions to obtain high quality data for 2 million CHO-K1 cells suspended in 0.5 mL of fully supplemented pre-warmed medium.

All datasets obtained from both dyes were seen to be positively skewed, showing that most of the data points are found below the median value of the distributions. Since the median is located quite close to the highest peak, which can be clearly observed through the

density/violin plots (particularly Figures 4.11, 4.12, and 4.14), it can be concluded that the Go/G1 population is the largest one. This was expected since cells were sampled according to conditions that were established by the experiments covering the cell culture process optimisation studies (see section 4.4). One of the conclusions of these experiments determined the sample collection point at 96 hours from the seeding point. The collection point is within the stationary phase; therefore, cells are not actively replicating as they do during the exponential phase. Positively skewed distributions are also observed in the data used for compensation analysis, the following section. However, this is not observed in data distributions obtained from the flow cytometric analysis of lectin sugar binding specificities as cells were collected during the exponential phase. As a consequence, data distributions are quite symmetric (see section 4.7).

To conclude, the experimental work intended to optimise the application of 7-AAD and DRAQ5 led to the determination of suitable concentration levels of these DNA dyes with the lowest level of data variability while providing sufficient fluorescence signal strength for detection.

4.6 Compensation analysis

An adequate experimental setup for a multicolour flow cytometric analysis ensures accurate and meaningful results. The presence of two or more fluorescent reagents on a single cell can lead to spillover. This phenomenon is characterized by the significant optical background a fluorescent reagent can cause to other reagents also present on a cell. Spillover occurs whenever the fluorescence emission of one dye is detected in a detector channel designed to measure signal from another dye. Since the extent of spillover is a linear function, the signal levels of the measured average can be corrected. This correction process is called compensation. Thus, compensation analysis is essential in order to properly visualise and analyse a complex dataset obtained from a multicolour flow cytometric experiment (Biosciences, 2009).

In this section, the level of brightness of samples individually stained with a lectin conjugated to V450 fluorochrome is evaluated. Such evaluation enables the identification of the lectin/V450 combination which produces the brightest signal. This combination is the most suitable V450 positive control for compensation analysis as it increases the accuracy level of a spillover matrix calculation (Biosciences, 2009). In turn, this matrix can then be used to compensate the data, removing the spillover, thereby determining the correct signal for each detector channel.

In an ascending order, Figure 4.15 shows a box plot of the data distributions measured through LECTIN-A channel. As an example, the unstained and stained samples were composed of cells collected from cultures in which the temperature was changed to 32°C in the last 24 hours. Since seven lectins compose the panel, seven datasets were collected from the stained samples, which were individually stained by a lectin/V450 combination. It can be observed that WGA/V450 combination is the brightest one, while LEC A/V450 is the dimmest. The median fluorescence signal obtained from cells stained with WGA/V450 was over three times as much as the signal measured from the unstained and LEC A/V450 samples.

The same analysis was performed for all the cell culture process variations, including the baseline conditions, and cells stained with V450 conjugated to WGA showed the brightest signal (this can be visualised later in the section 4.9 measured by LECTIN detector channels. Therefore, the WGA single-stained control was used for compensation analysis.

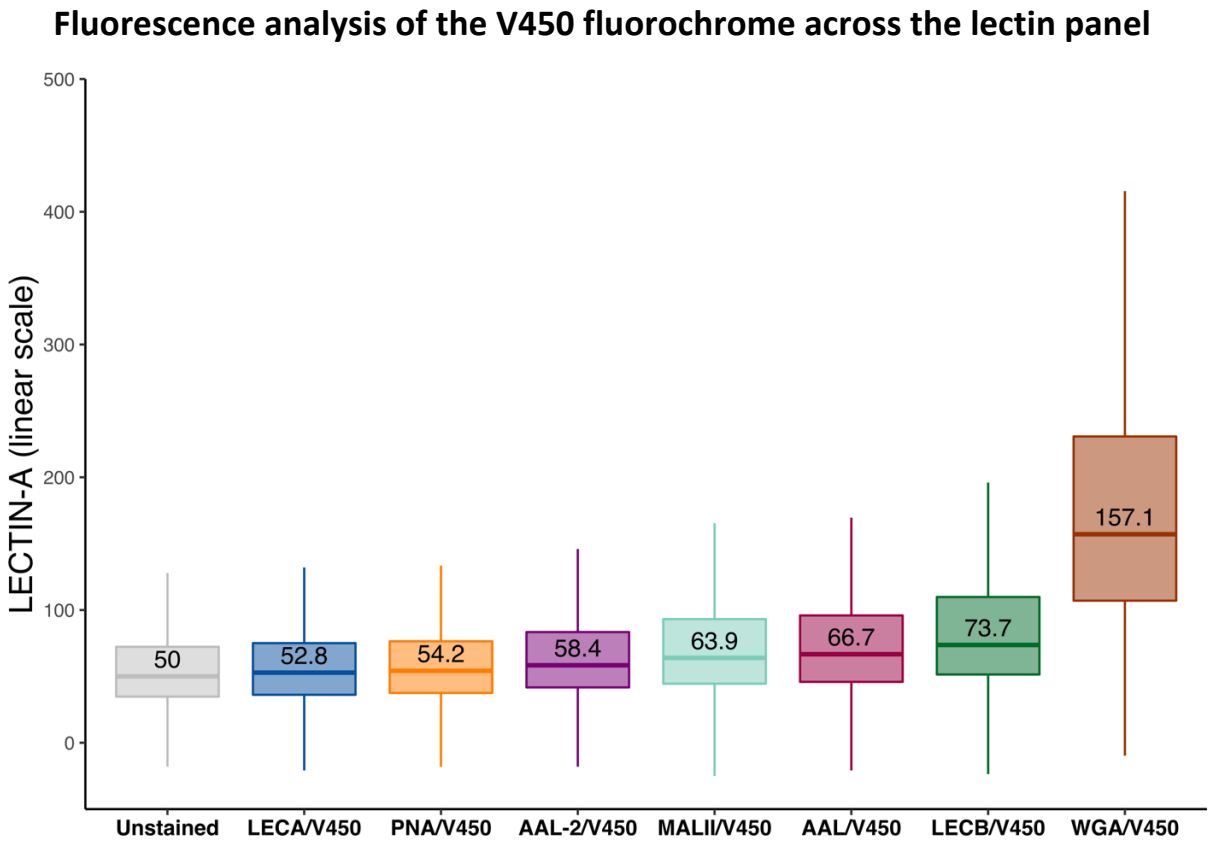

Figure 4.15: Box plot showing LECTIN-A data distributions of the unstained and seven lectin/V450 stained samples. These samples were collected from cultures subjected to a decrease in temperature from 37 to 32°C in the last 24 hours of culture. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.4.2).

After the determination of the brightest positive control for V450, the full evaluation of the spillover was then performed by calculating the spillover matrix. This calculation requires single-stained controls for all the fluorescent reagents of a polychromatic flow cytometric analysis. 7-AAD and DRAQ5 single-stained controls, thus, were used in conjunction with the V450 control. With the aid of the *flowCore* R package, the matrix for 32°C experiment was then calculated (Figure 4.16).

The matrix in Figure 4.16 shows that 7-AAD spills into the DRAQ5-A channel (as much as 22% of 7-AAD fluorescence in its own channel). V450 spills 69% into 7-AAD-A and 7% into DRAQ5-A. In contrast, DRAQ5 fluorescence is only detected on its own channel.

	7-AAD-A	DRAQ5-A	LECTIN-A
7-AAD	1	0.22	0
DRAQ5	0	1	0
V450/WGA	0.69	0.07	1

Figure 4.16: Spillover matrix calculated from 32°C experiment compensation controls. The matrix shows the amount of spillover of the fluorescent reagents into another detector channels.

The mathematical inversion of the spillover matrix generates the compensation matrix, which in turn was used to calculate the true signal of a channel (Biosciences, 2009). Figure 4.17 shows the inverted matrix and Equations 4.1 to 4.3 show the calculation of compensated values for each channel.

	7-AAD-A	DRAQ5-A	LECTIN-A
7-AAD	1	-0.22	0
DRAQ5	0	1	0
V450/WGA	-0.69	0.081	1

Figure 4.17: Compensation matrix calculated by mathematically inverting the spillover matrix from 32°C experiment.

The compensated value of a channel is equal to the signal measured through this channel (values in percentage calculated in the spillover matrix) multiplied by the summation of the appropriate coefficients of the compensation matrix:

$$\mathbf{7-AAD-A_{TRUE}} = 100 \times (1 + 0 - 0.69) = 31 \qquad \text{Equation } (\mathbf{8})$$

$$\mathbf{DRAQ5 - A_{TRUE}} = 100 \times (-0.22 + 1 + 0.081) = 86 \qquad \text{Equation } (\mathbf{9})$$

$$\mathbf{V450_{TRUE}} = 100 \times (0 + 0 + 1) = 100 \qquad \text{Equation } (\mathbf{10})$$

Since fluorescence values were expected to change due to cells experiencing different growing conditions, compensation controls were prepared and analysed for each experimental run. Thus, the compensation analysis described in this section was conducted for all the experiments involving the use of more than two fluorescent reagents.

Therefore, compensation analysis was successfully performed allowing the accurate analysis of datasets obtained from multicolour flow cytometric experiments.

4.7 Flow cytometric analysis of lectin sugar binding specificity

Previously, in the section 4.3 sugar binding specificities of AAL-2, LEC A and LEC B were determined. However, since these lectins were intended for use as probes to glycoprofile the surface of cells using flow cytometry, a flow cytometric experiment was set up to analyse the sugar binding specificities of the lectin panel: AAL-2, LEC A, LEC B, AAL, MAL II, PNA, and WGA. The results of this experiment allowed the determination of lectin specificity within the conditions in which the lectins were used for probing CHO-K1 cell surface.

Firstly, this section evaluates the potential binding of V450 streptavidin to the cell in the absence of a biotinylated lectin. Then, the analysis of sugar binding specificity of each lectin is explored.

The binding of V450 streptavidin fluorochrome is analysed in Figure 4.18 which shows a box plot with data distributions of signal measured through LECTIN-A detector channel. Apart from the cells used for the unstained sample, 7-AAD was added to the other two samples to distinguish dead cells. As a result, the unstained sample has demonstrated to possess higher variability which is shown by the distance between the lower and upper quartiles for instance.

Such variability could be mostly due to the inclusion of data from dead cells, which were unable to be removed as 7-AAD was not added to this sample.

As has been shown in the previous section, 7-AAD single-stained sample does not affect the signal measured by LECTIN-A channel. In fact, the signal detected from this fluorochrome is lower than the unstained sample as the median values revealed, 51.4 for the former and 57 for the latter.

Similarly, the sample stained with both 7-AAD and V450 has demonstrated a slightly lower median value (55.6) than the unstained sample. Consequently, it can be concluded that V450 streptavidin reagent does not interact with the cells in the absence of biotinylated biomolecules such as the lectins used as probes for the analysis of cell surface glycoprofile.

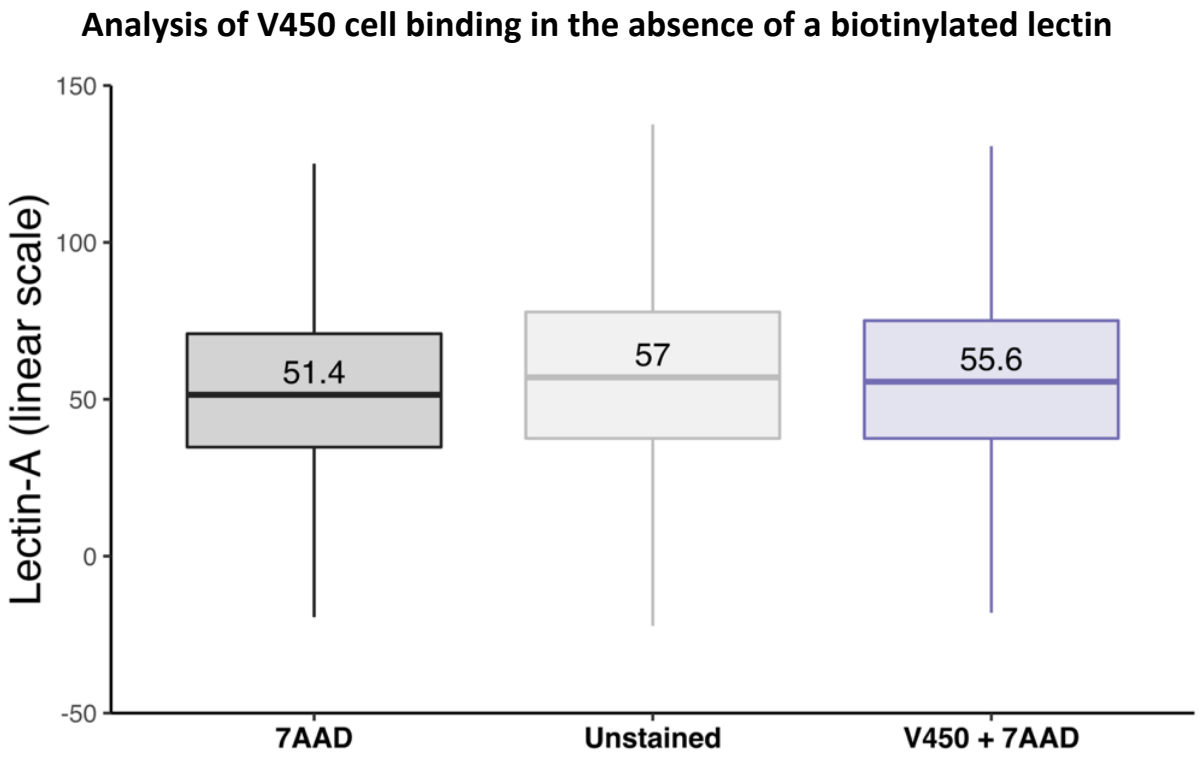

Figure 4.18: Box plot showing LECTIN-A data distributions of unstained, 7-AAD and V450 + 7-AAD stained samples. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.7).

The sugar binding specificity of each lectin was then analysed. Thus, some samples were prepared with lectins which were previously incubated with free sugar molecules with the purpose of investigating the inhibition of lectin cell binding owing to the interaction with a free sugar. The methodology used is described in more detail in section 3.12.1.7. Since 7-AAD was used to look at only live cells from all samples used for lectin specificity investigation, the 7-AAD single-stained sample was then used as the negative control for V450/lectin. Therefore, the *unstained* sample in the seven following plots are in fact unstained for V450/lectin but stained for 7-AAD.

Figure 4.19 shows a box plot demonstrating the LECTIN-A data distributions of unstained and AAL samples. The signal measured from samples in which AAL lectin was previously incubated with L-Fucose and Mannose, both demonstrated to substantially reduce the level of lectin binding to the cell. However, L-Fucose inhibited AAL cell binding to a slightly higher extent than Mannose. Although the ELLA analysis results involving AAL sugar binding specificity (Figures 4.5 and 4.6) showed that the lectin affinity for binding Fucose was considerably stronger than for Mannose, AAL also demonstrated higher affinity for Fucose in the flow cytometric analysis.

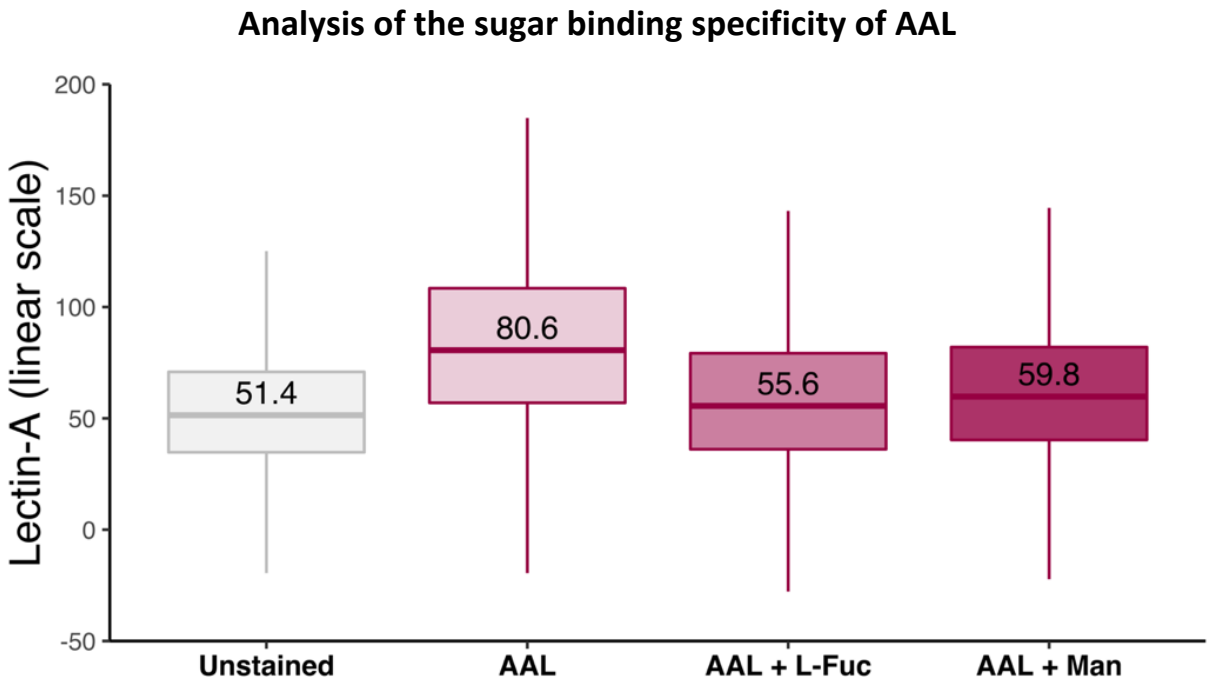

Figure 4.19: Box plot showing the fluorescence signal data distributions (measured through LECTIN-A channel) of an unstained sample and AAL stained samples. AAL + L-Fuc and AAL + Man are samples in which AAL was previously incubated with L-Fucose and Mannose free sugar molecules. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.7).

Figure 4.20 shows a box plot demonstrating the LECTIN-A data distributions of unstained and LEC A samples. The signal measured from the sample in which LEC A lectin was previously incubated with Galactose demonstrated to reduce the level of lectin binding to the cell. The median value of this distribution is equal to the unstained sample, 51.4. However, the variability of LEC A + Gal sample is slightly higher than the unstained sample variability as can be seen by the distance between the upper and lower quartiles for instance. LEC A + Man sample, in contrast, demonstrated no inhibition effect on LEC A cell binding, in fact, Mannose seemed to improve the binding of LEC A to the cell since the median value of this sample distribution was 59.8 in comparison to 52.8 of LEC A sample.

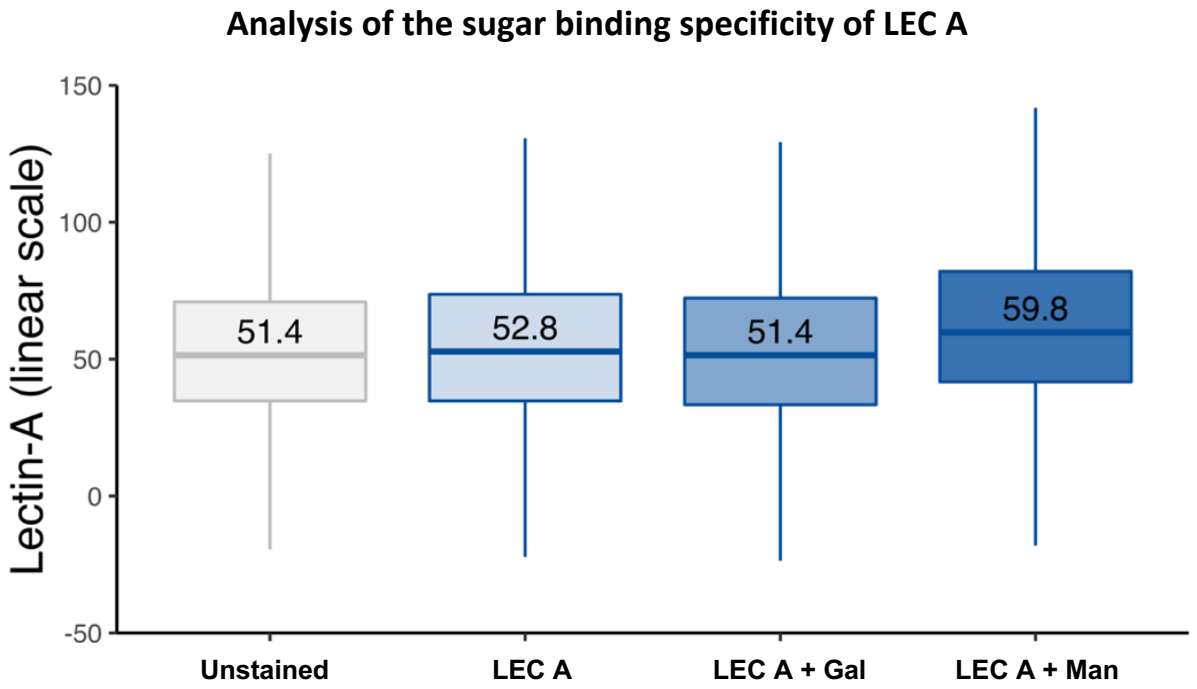

Figure 4.20: Box plot showing the fluorescence signal data distributions (measured through LECTIN-A channel) of an unstained sample and LEC A stained samples. LEC A+ Gal and LEC A + Man are samples in which LEC A was previously incubated with Galactose and Mannose free sugar molecules, respectively. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.7).

A box plot demonstrating the LECTIN-A data distributions of unstained and LEC B samples is shown in Figure 4.21. LEC B + Galactose and LEC B + L-Fucose samples both demonstrated a reduction in the fluorescence median values in relation to LEC B sample, 55.6, 52.8 and 58.4, respectively. However, L-Fucose inhibited LEC B from binding to the cells to a greater extent, reducing the fluorescence signal close to the unstained median value which was 51.4. This result is in agreement with the ELLA analysis which showed a stronger LEC B affinity for Fucose in relation to Galactose (Figures 4.5 and 4.6). Although the ELLA analysis for LEC B also showed that Mannose was the strongest interaction with the lectin in comparison to Galactose and Fucose, LEC B + Mannose sample demonstrated no inhibition effect on the lectin. LEC B + Mannose data distribution is quite similar to LEC B distribution as the median value is the same and data variability shows great similarity as well. Also, LEC B + GlcNAc sample demonstrated

no effect on LEC B inhibition. In fact, it showed a slightly improved effect in the lectin cell binding, since the median value calculated for the sample was 62.5.

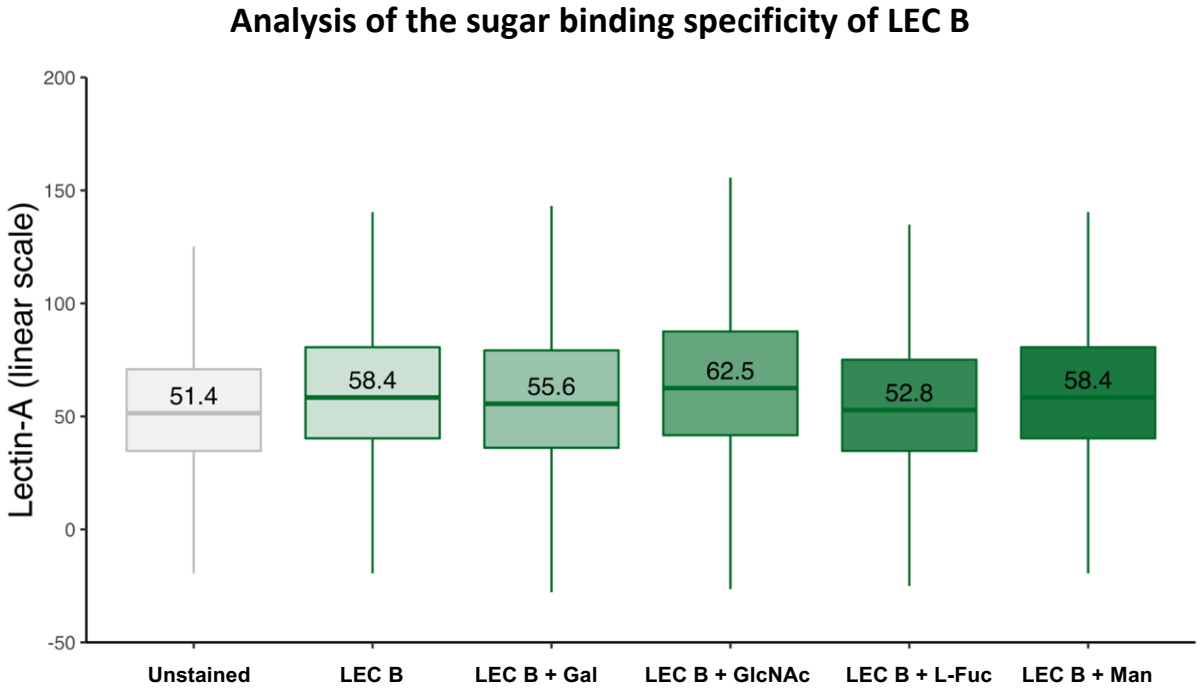

Figure 4.21: Box plot showing the fluorescence signal data distributions (measured through LECTIN-A channel) of an unstained sample and LEC B stained samples. LEC B + Gal, LEC B + GlcNAc, LEC B + L-Fuc, and LEC B + Man are samples in which LEC B was previously incubated with Galactose, N-Acetylglucosamine, L-Fucose and Mannose free sugar molecules, respectively. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.7).

Figure 4.22 shows a box plot demonstrating the LECTIN-A data distributions of unstained and PNA samples. The signal measured from the sample in which PNA lectin was previously incubated with Galactose demonstrated a reduction in the level of lectin binding to the cell. PNA + Gal data distribution showed a median value equal to the unstained sample, 51.4. Conversely, PNA + Man sample demonstrated no inhibition effect on the PNA cell binding. In fact, the incubation of PNA with free Mannose molecules enhanced the lectin cell binding, as the median value measure for PNA + Man sample was 57 while the PNA sample value was 52.8.

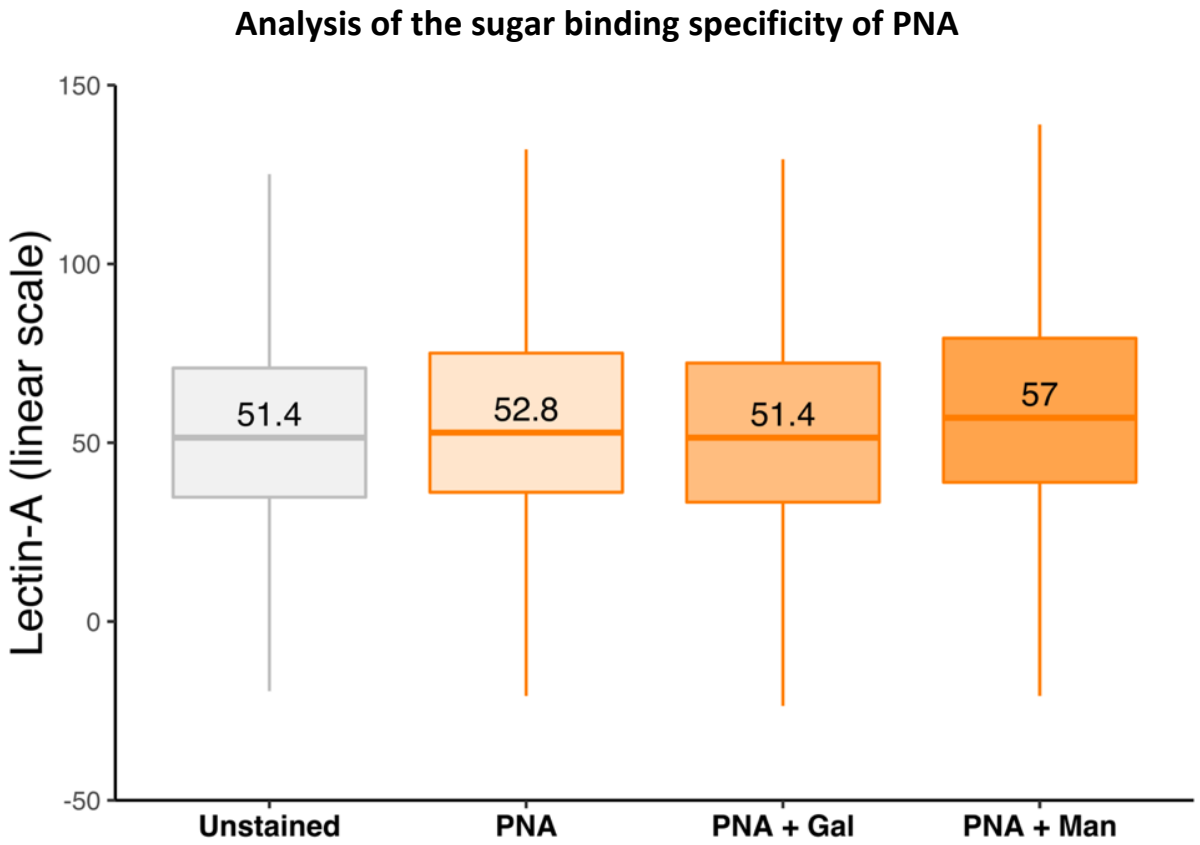

Figure 4.22: Box plot showing the fluorescence signal data distributions (measured through LECTIN-A channel) of an unstained sample and PNA stained samples. PNA + Gal, PNA + Man are samples in which LEC B was previously incubated with Galactose and Mannose free sugar molecules, respectively. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.7).

Figure 4.23 shows a box plot demonstrating the LECTIN-A data distributions of unstained and AAL-2 samples. The signal measured from samples in which AAL-2 lectin was previously incubated with N-Acetylgalactosamine (GalNAc) and N-Acetylglucosamine (GlcNAc), both demonstrated to substantially reduce the level of lectin binding to the cell. However, GlcNAc inhibited AAL-2 cell binding to a higher degree than GalNAc, reducing the data distribution media value to 52.8 which is quite close to the unstained median value of 51.4. ELLA analysis involving AAL-2 sugar binding specificity (Figure 4.5) also detected the affinity of the lectin for GlcNAc.

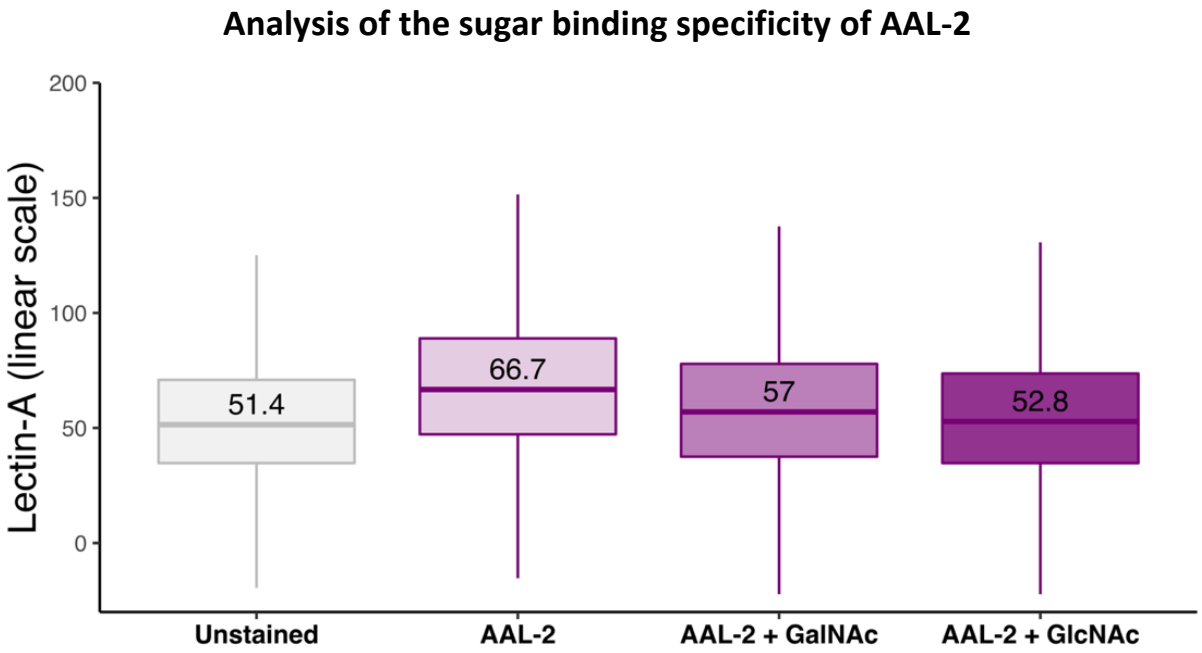

Figure 4.23: Box plot showing the fluorescence signal data distributions (measured through LECTIN-A channel) of an unstained sample and AAL-2 stained samples. AAL-2 + GalNAc, AAL-2 + GlcNAc are samples in which AAL-2 was previously incubated with N-Acetylgalactosamine and N-Acetylglucosamine, respectively. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.7).

A box plot demonstrating the LECTIN-A data distributions of unstained and MAL II samples can be seen in Figure 4.24. The fluorescence signal measured from MAL II + GlcNAc sample demonstrated no inhibition effect on MAL II cell binding. This is clearly seen in the sample median value of 72.3 which is equal to MAL II sample value. However, MAL II + GlcNAc showed a higher level of data variability in comparison to MAL II sample as can be observed, for instance, in the distance between the extremes of its vertical bar, which shows the range from the smallest to the largest nonoutlying values. The higher level of data variability of MAL II + GlcNAc in relation to MAL II sample can also be observed in the distance between the upper and lower quartiles (box's height). MAL II + sialic acid showed strong inhibition effect on MAL II cell binding since the median value was 51.4 which was equal to the unstained median value. However, MAL II + sialic acid data variability was slightly higher than the unstained one.

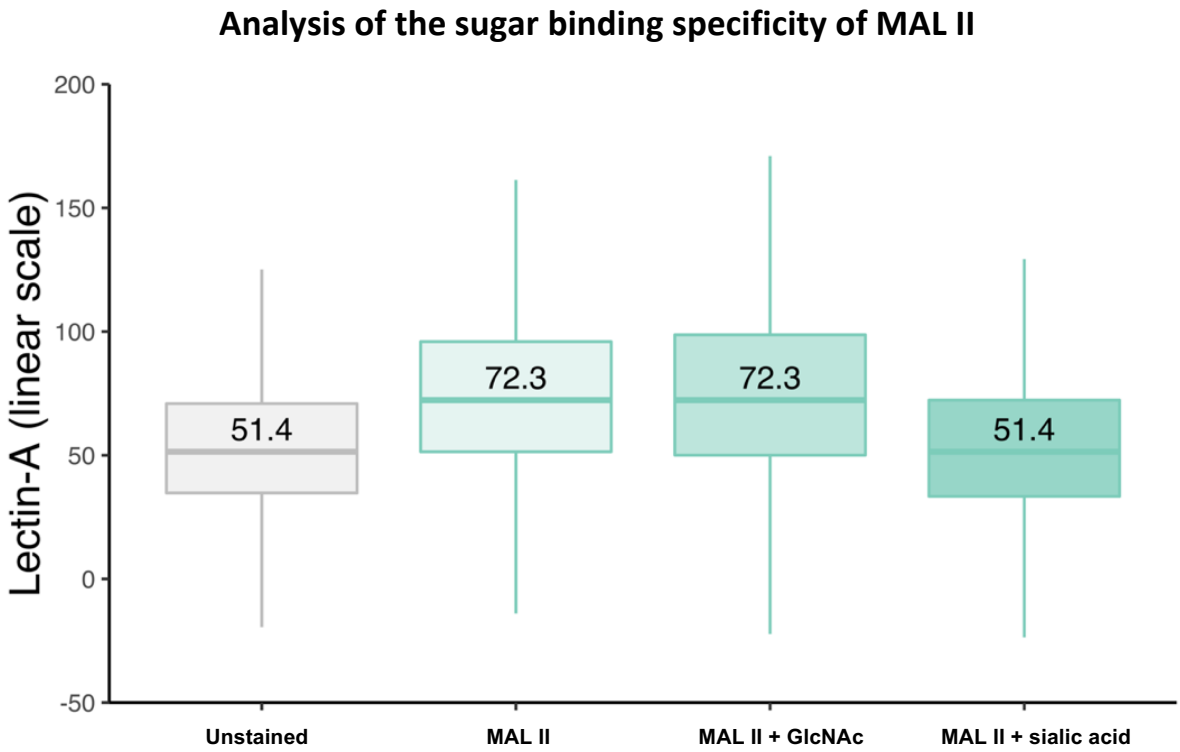

Figure 4.24: Box plot showing the fluorescence signal data distributions (measured through LECTIN-A channel) of an unstained sample and MAL II stained samples. MAL II + GlcNAc, MAL II + sialic acid are samples in which MAL II was previously incubated with N-Acetylglucosamine and Sialic acid, respectively. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.7).

Figure 4.25 shows a box plot demonstrating the LECTIN-A data distributions of unstained and WGA samples. The signal measured from samples in which WGA lectin was previously incubated with N-Acetylgalactosamine (GalNAc) and N-Acetylglucosamine (GlcNAc), both demonstrated to reduce the level of lectin binding to the cell to a great extent. However, GlcNAc inhibited WGA cell binding to a higher degree than GalNAc reducing the data distribution media value to 50 which is even lower than the unstained median value of 51.4. ELLA analysis involving WGA sugar binding specificity (Figure 4.5) also detected the affinity of the lectin for GlcNAc.

To summarise the outcomes of the last 7 plots, Table 4.3 shows the sugar binding specificities of each lectin. The table allows the observation of binding similarities between in-house

production lectins and commercial ones. LEC A and PNA both demonstrated affinity for Galactose, AAL-2 and WGA demonstrated affinity for N-Acetylglucosamine and N-Acetylgalactosamine, both showing stronger binding specificity for N-Acetylglucosamine. Whereas AAL and LEC B, demonstrated a strong affinity for L-Fucose. Additionally, the binding specificity results of commercial lectins are in agreement with the information provided by the supplier (VectorLabs).

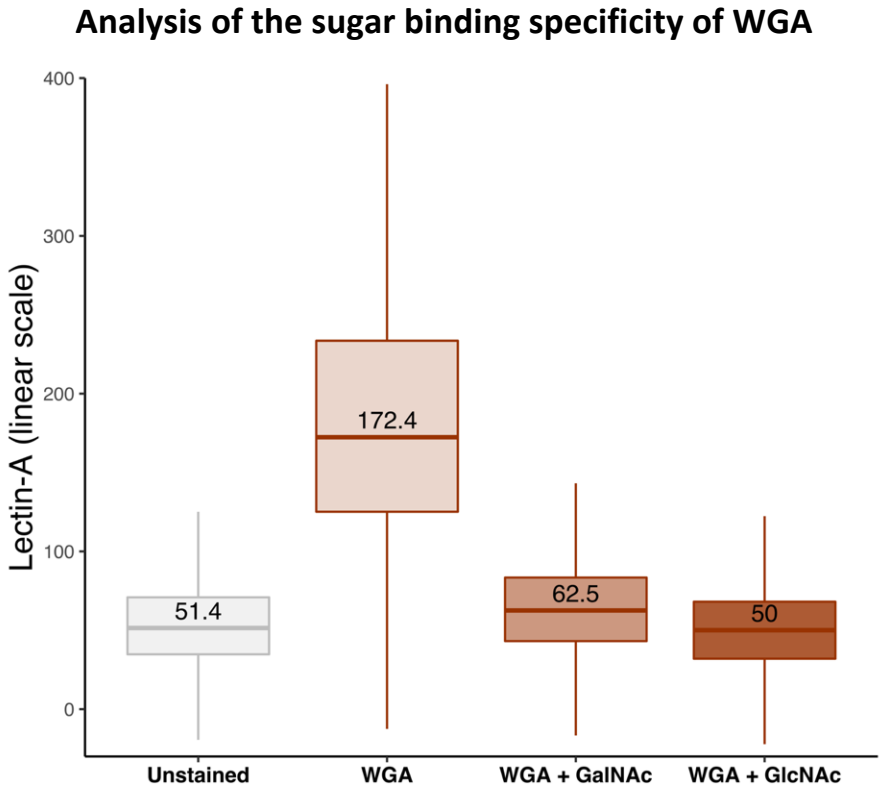

Figure 4.25: Box plot showing the fluorescence signal data distributions (measured through LECTIN-A channel) of an unstained sample and WGA stained samples. WGA + GalNAc, WGA + GlcNAc are samples in which WGA was previously incubated with N-Acetylgalactosamine and N-Acetylglucosamine, respectively. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.7).

Table 4.3: Table summarising the flow cytometric results of sugar binding specificity of lectins.

Lectin	Strongest sugar-binding molecule	2nd strongest sugar-binding molecule
AAL	L-Fucose	Mannose
LEC A	Galactose	-
LEC B	L-Fucose	Galactose
PNA	Galactose	-
AAL-2	N-Acetylglucosamine	N-Acetylgalactosamine
MAL II	Sialic acid	-
WGA	N-Acetylglucosamine	N-Acetylgalactosamine

Although AAL and LEC B have shown results which disagree with the ELLA analysis to a certain degree (see section 4.3), this fact does not affect the purpose of the experimental work covered by this present section, which is to characterize lectin sugar binding within the flow cytometric experimental conditions. The differences in the outcomes between the ELLA and flow cytometry analyses could have been caused by a number of differing conditions which might have acted individually or in combination with each other. The concentration levels of the glycoproteins and free sugars and the solutions used (TBS is used in ELLA, while the cell culture media is the solution of choice in the flow cytometric assay) are examples of these differing conditions. Furthermore, immobilized glycoproteins bearing specific carbohydrate residues were used in the ELLA analysis, whereas free sugar molecules were used in the flow cytometric study. Above all, due to the strong quantitative nature of flow cytometry, this technique is more capable of demonstrating biological phenomena involving the measurement of binding activities through the use of fluorescent reagents.

To conclude, the flow cytometric analysis of sugar binding specificities of lectins has successfully demonstrated that free sugar molecules when incubated with lectins prior to probing the cells, can prevent the lectins from interacting with sugar molecules on cell surface. Most importantly, the analysis has allowed the determination of sugar binding specificities of each lectin while providing information on their biological activity status within the flow

cytometric experimental conditions. Since all lectins have shown cell binding inhibition due to interaction with a particular free sugar molecule, which demonstrates their biological activity, the panel of lectins has proved suitability for cell surface glycoprofiling by flow cytometric analysis.

4.8 Lectin cytotoxicity: a flow cytometry-based analysis

The phenomenon of lectin cytotoxicity can occur precisely after lectins have bound to all binding sites available on the cell surface, leaving the unbound lectins outside the cell in a highly concentrated level in relation to the internal environment of the cell (Stanley and Sundaram, 2014).

In order to determine the non-toxic concentration level of lectins to probe CHO-K1 cells, flow cytometric cytotoxicity studies were performed. Lectin concentration levels were varied from 0 to 12.5 μg/mL and cell viability was determined for each variation (see section 3.12.1.6). Since the previous section has indicated sugar binding specificity similarities between in-house production lectins and commercial ones, this section presents the lectin cytotoxicities in pairs: AAL & LEC B, PNA & LEC A, WGA & AAL-2. However, MAL II is evaluated individually. In addition, the scale of the viability axis of the following four plots is fixed; that is, it ranges from 96% to 100%, and the dimensions of the figures are identical. This allows the direct comparison of data variability among the lectin panel. Such variability is demonstrated by the 95% confidence interval of the polynomial fit of each lectin dataset.

Figure 4.26 shows a line plot with the polynomial fits of experimental data from AAL and LEC B cytotoxicity studies. In addition, the plot allows the observation of 95% confidence interval.

A polynomial fit of the data obtained when using AAL as probe revealed a narrower 95% confidence interval than when LEC B was used. Consequently, AAL data variability was shown to be lower than LEC B. Cells incubated with increasing concentrations of the in-house production lectin, LEC B, showed viability around the 97% up to 6.25 μg/mL, but demonstrated a tendency to sharply decrease in cell viability as the concentration was raised beyond 6.25

μg/mL. In contrast, the commercial lectin, AAL, was shown to be slightly less toxic to the cells since the cell viability is constantly around 99% across the entire range of lectin concentration. However, up to 12.5 μg/mL, both lectins demonstrated to be non-toxic to the cells as viability is well above 90%.

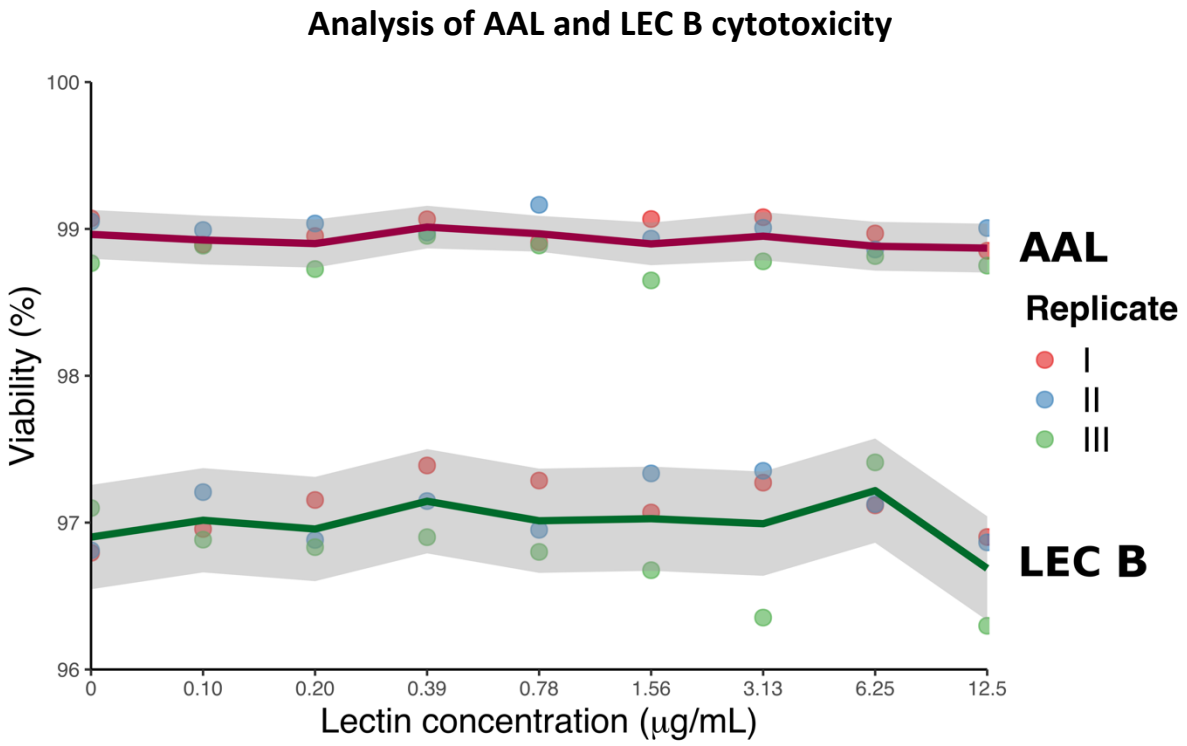

Figure 4.26: Line plot showing AAL and LEC B cytotoxicity polynomial fits along with 95% confidence intervals. Data obtained through a BD FACSAria™ I flow cytometer and using 7-AAD to stain dead cells, thus determining cell viability (see section 3.12.1.6).

PNA and LEC A cytotoxicity data is shown in Figure 4.27. Although the in-house production lectin, LEC A, gave a narrower 95% confidence interval than PNA, LEC A demonstrated to be slightly more toxic to the cells than PNA. While PNA viability values mostly fell between 98% and 99%, LEC A values fell between 97.5% and 98%. Nevertheless, both lectins revealed to be non-toxic to the cells up to 12.5 μg/mL.

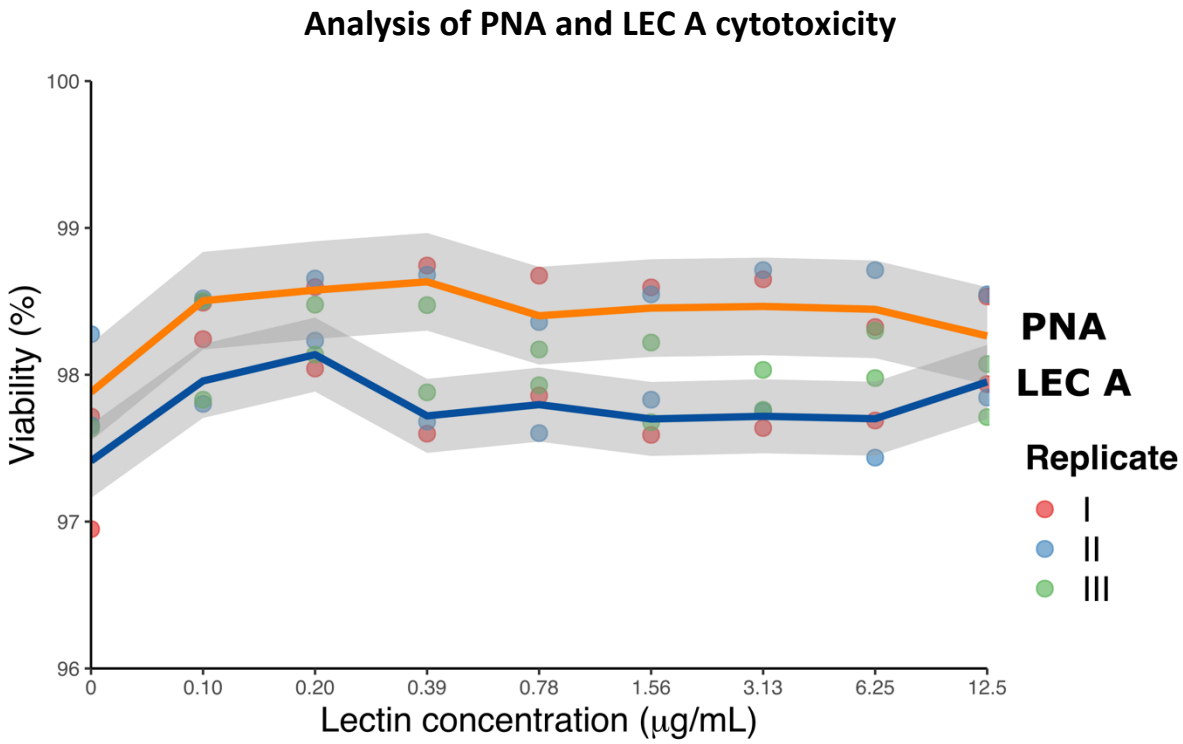

Figure 4.27: Line plot showing PNA and LEC A cytotoxicity polynomial fits along with 95% confidence intervals. Data obtained through a BD FACSAria™ I flow cytometer and using 7-AAD to stain dead cells, thus determining cell viability (see section 3.12.1.6).

Figure 4.28 demonstrates WGA and AAL-2 cytotoxicity data. It can be observed that both lectins revealed the same level of data variability as 95% confidence interval widths are quite similar across the entire concentration range. In addition, the lectins showed a tendency of a sharp drop in cell viability as lectin concentration was increased beyond 6.25 μg/mL. However, likewise LEC B and LEC A, AAL-2 demonstrated to be slightly more toxic to the cells than its equivalent commercial lectin, WGA. While WGA viability values were mostly between 99% and 100%, AAL-2 values fell mostly between 98% and 99%. As a result, both lectins are very safe to be used on the cells up to the concentration level of 12.5 μg/mL.

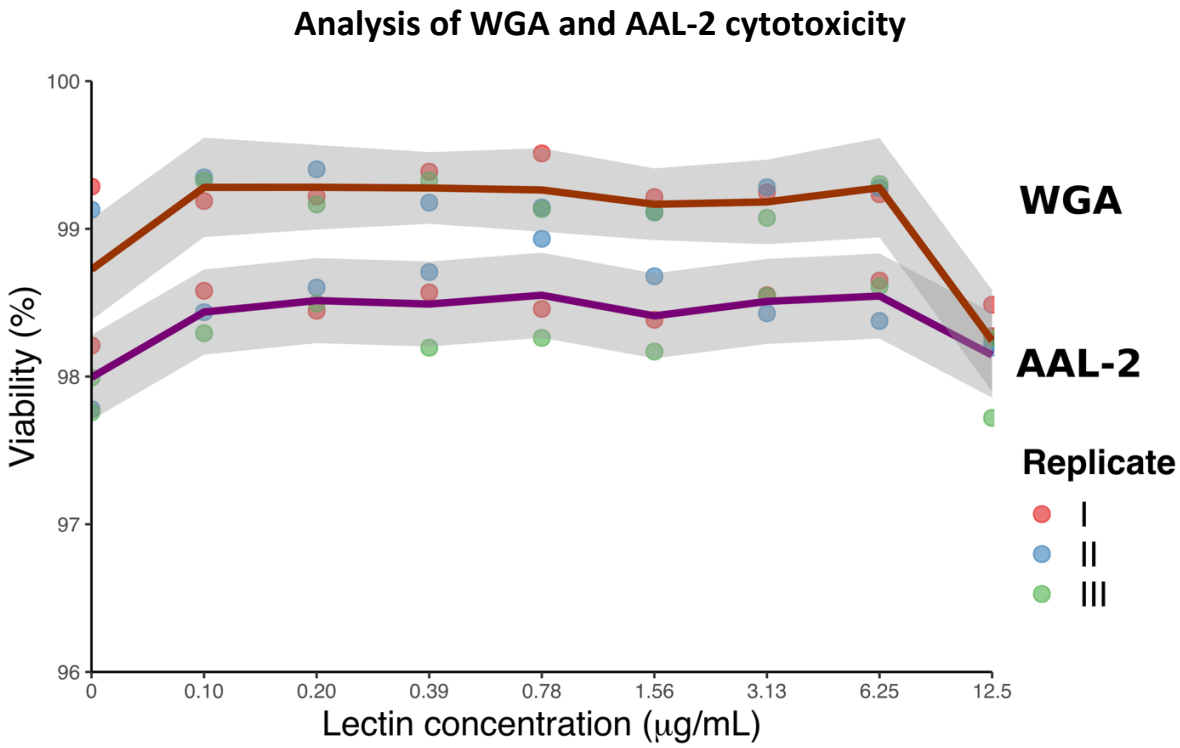

Figure 4.28: Line plot showing WGA and AAL-2 cytotoxicity polynomial fits along with 95% confidence intervals. Data obtained through a BD FACSAria™ I flow cytometer and using 7-AAD to stain dead cells, thus determining cell viability (see section 3.12.1.6).

And finally, Figure 4.29 demonstrates MAL II cytotoxicity data. The viability values of cells incubated with increasing concentration levels of MAL II fell mostly between 98% and 99%. However, the data revealed a tendency to drop as cells were exposed to concentration levels higher than 1.56 μg/mL. Furthermore, MAL II as well as LEC B data variability was shown to be the highest among the lectin panel, while AAL revealed the lowest variability as can be observed from the comparison of the 95% confidence intervals. In conclusion, MAL II cytotoxicity study showed that the lectin is non-toxic to the cells up to 12.5 μg/mL.

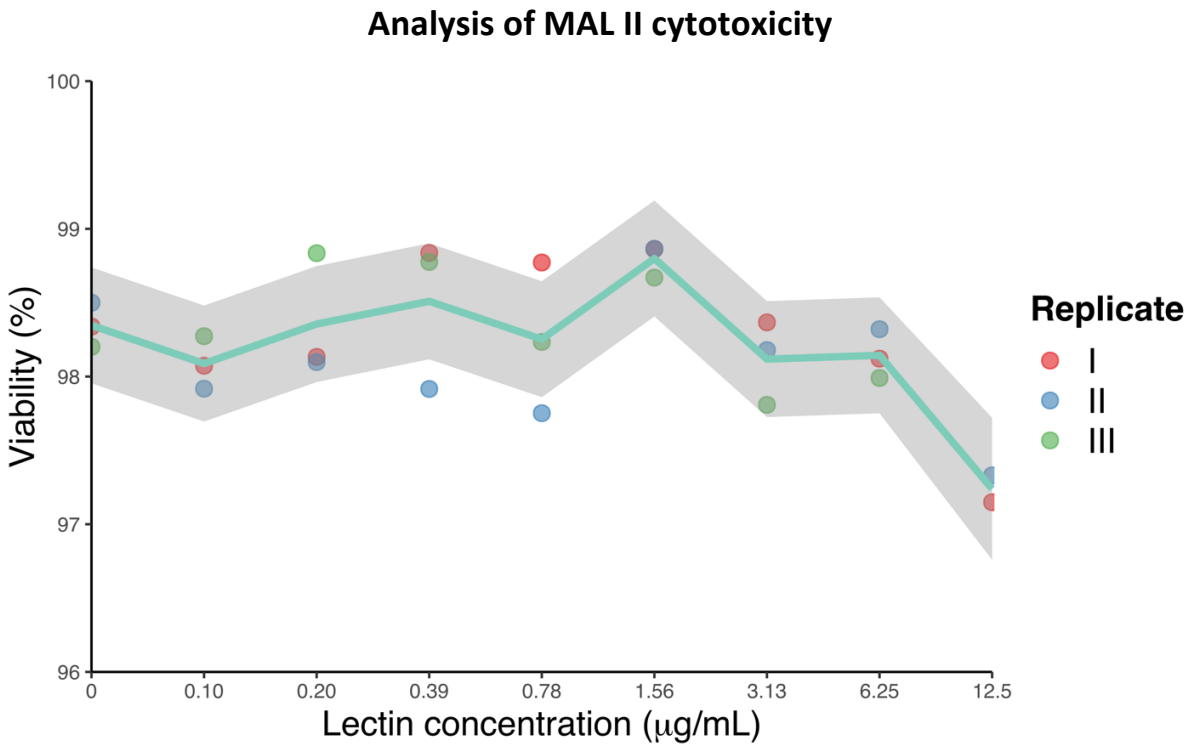

Figure 4.29: Line plot showing MAL II cytotoxicity polynomial fit along with 95% confidence interval. Data obtained through a BD FACSAria™ I flow cytometer and using 7-AAD to stain dead cells, thus determining cell viability (see section 3.12.1.6).

In the form of a line plot, Figure 4.30 summarises the polynomial fits of all lectins investigated in the cytotoxicity studies. However, 95% confidence interval is not shown in the figure for better visualization of the trends revealed by each lectin.

AAL has demonstrated to be the most stable curve in contrast to MAL II, the least one. The plot allows the observation of the two safest lectins to be used to probe CHO-K1 cells as well as the two least ones. WGA and AAL are the safest with cell viability values around 99% while LEC B and LEC A are the least safe ones with viability values around 97% and 97.5%, respectively. Although such conclusion can be made, it is important to stress the fact that all lectins of the panel have not shown a severe or even a mild toxicity effect since the viability values range from 96.5% to 99.5% up to 12.5 μg/mL.

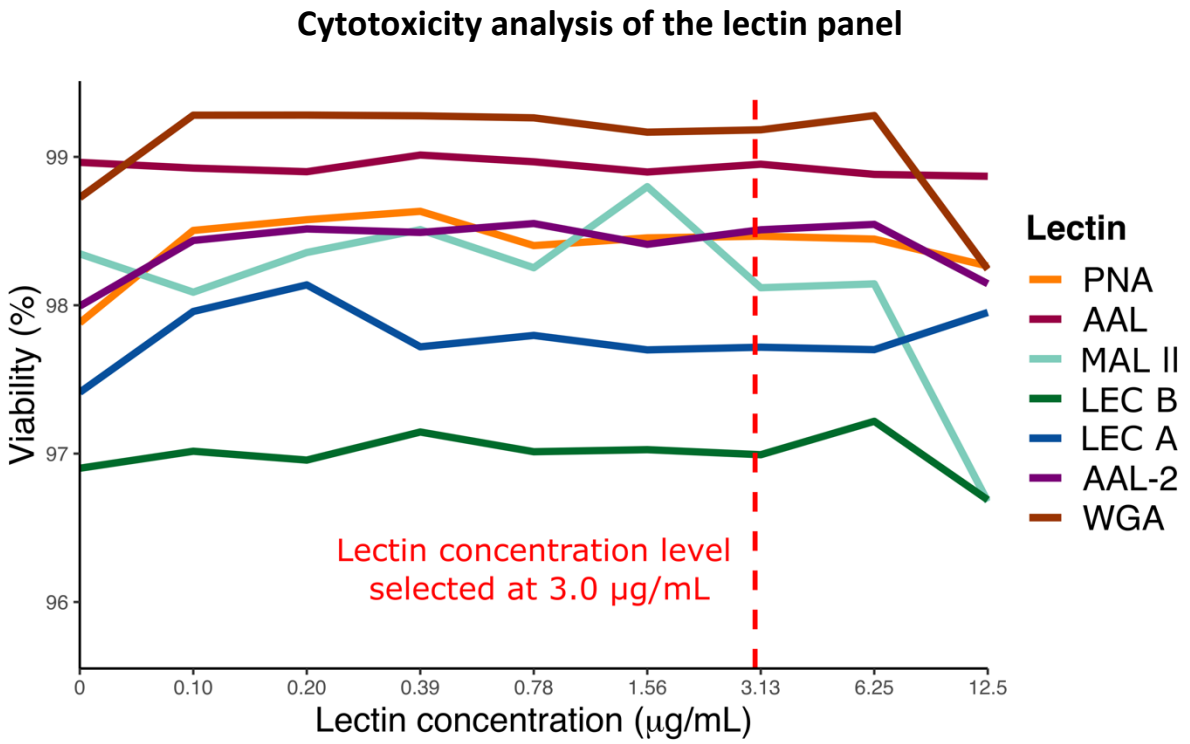

Figure 4.30: Line plot showing polynomial fits of all lectins involved in the flow cytometric cytotoxicity studies. The red dashed line shows the concentration level of lectin selected for the investigation of the variation of cell surface glycoprofile. Data obtained through a BD FACSAria™ I flow cytometer and using 7-AAD to stain dead cells, thus determining cell viability (see section 3.12.1.6).

Furthermore, the plot allows the observation of the concentration level at which most lectins cause a drop in the cell viability. Therefore, the determination of the lectin concentration level to use for the investigation of the variation of cell surface glycoprofile could be done. As can be seen in the plot, although 6.25 μg/mL is the concentration level at which most lectins cause a decrease in viability, the concentration level of choice was 3.0 μg/mL as illustrated by the red dashed line. Such lectin concentration level is within a range in which viability levels are stable for all lectins except for MAL II. Additionally, a higher lectin concentration level requires a great increase in the V450 volume for the preparation of lectins for cell probing (see section 3.12.1.6), since the lectin volume is directly proportional to the V450 volume. Consequently,

the costs associated with commercial lectins and V450 purchasing would have escalated which could have compromised the research project budget.

In sum, the outcomes of lectin cytotoxicity studies have successfully allowed the determination of a financially suitable and quite safe lectin concentration level for the investigation of the glycoprofile variation of CHO-K1 cell surface.

4.9 Variation of cell culture parameters: Statistical analysis of the effects on CHO-K1 cells

Although the achievement of the results presented in previous sections allowed the development of an optimised flow cytometric methodology, the application of this methodology to investigate the effects of the alteration of cell culture parameters (level of spent medium, temperature and CO_2) on the cell surface glycoprofile, generated about 300 million flow cytometric observations in which 15 parameters (FSC-A, SSC-A, 7-AAD-A, and LECTIN-A for instance) were measured for each observation. Based on these parameters a number of variables was determined enabling the classification of single cells in terms of their DNA cycle stage and viability for instance. The process of removing datapoints extracted from cell debris and aggregates reduced the dataset to 80 million single cell observations (see section 3.12.5).

Therefore, this section presents the statistical analysis and discussion of the data obtained from the experiments involving the variation of cell culture parameters (see section 3.12.2). For each cell culture parameter variation, the effects on pH and viability are presented and discussed first. Then descriptive statistical analysis and discussion of the effects on the relative cell size and the cell internal and external complexity level is developed. An in-depth statistical analysis of cell surface glycoprofile variation is shown and the results of the descriptive and inferential analysis are discussed. Furthermore, the results of power analysis are then discussed in detail in order to evaluate whether the findings are scientifically meaningful.

Lastly, a general comparative power analysis is developed to identify the cell culture parameter and the lectins/glycans associated with the most scientifically meaningful results.

The results are summarised in the form of line plots, box plots, bar plots and tables. However, the variability of the data is not shown in line plots to facilitate the visualisation of the trends revealed by the curves, except for line plots illustrating the pH variation as they contain only one curve each.

4.9.1 Analysis of the effects of spent medium level variation

The experimental data involving the variation of spent medium levels comprises 7 datasets each representing a spent medium level: 6 datasets consist of the variation of the spent medium levels in the last 24 hours of culture, and 1 dataset extracted from cells grown under the baseline condition throughout the entire 96 hours of culture (sample collection point).

Several components change their concentrations as the levels of a spent medium is varied such as an increase in alkaline compounds accompanied with a decrease in nutrient levels in spent media used for longer periods of time. Although the overall variation of the cell culture medium composition can affect the glycosylation process, the discussion of the results of this section is focused on the variation of the levels of nutrients, since this variation can greatly affect the availability of the building blocks needed to assemble different and multiple carbohydrate structures in the cells.

In order to facilitate the understanding of the data, the variables of interest were plotted against the variation of spent medium levels in relation to the setpoint or the baseline condition. In other words, the 0 point of the x axis (Level of spent medium) depicts the data obtained from cells cultivated under baseline conditions, whereas the remaining points depict the datasets obtained from the different levels of spent medium which were measured as a day unit. Negative values in the x axis consist of media with lower (depleted) levels of nutrients in relation to the 0 point (baseline condition), while the positive values consist of media with

higher (excess) levels of nutrients. The experimental setup is fully described in section 3.12.2.2.

Cell culture parameters

Viability and pH parameters were measured to characterize the cell culture process involving the variation of nutrient levels in the medium. While viability was obtained through flow cytometry with the aid of 7-AAD fluorochrome (see sections 1.5, 3.12.1.2, and 3.12.1.3), pH was measured using a pH electrode (see section 3.12.1.8).

It can be observed through Figure 4.31 that cell viability for all lectin samples significantly increased between -3 and -1 day as spent medium levels increased as well. Cell viability continued to rise between -1 and +1 day; however, the rate of this increase was lower than the initial one and the cell viability maximum reaching point was at 99%. Further increases in the spent medium levels allowed cell viability to stabilise at 99%.

The low cell viability values (the lowest value was 88% for WGA samples) between -3 and -1 day showed that cells were subjected to spent medium levels which significantly affected their growth. However, between +1 and +3 days, cells were treated with extra levels of nutrients in relation to the baseline point, thereby allowing the cells to continue to grow.

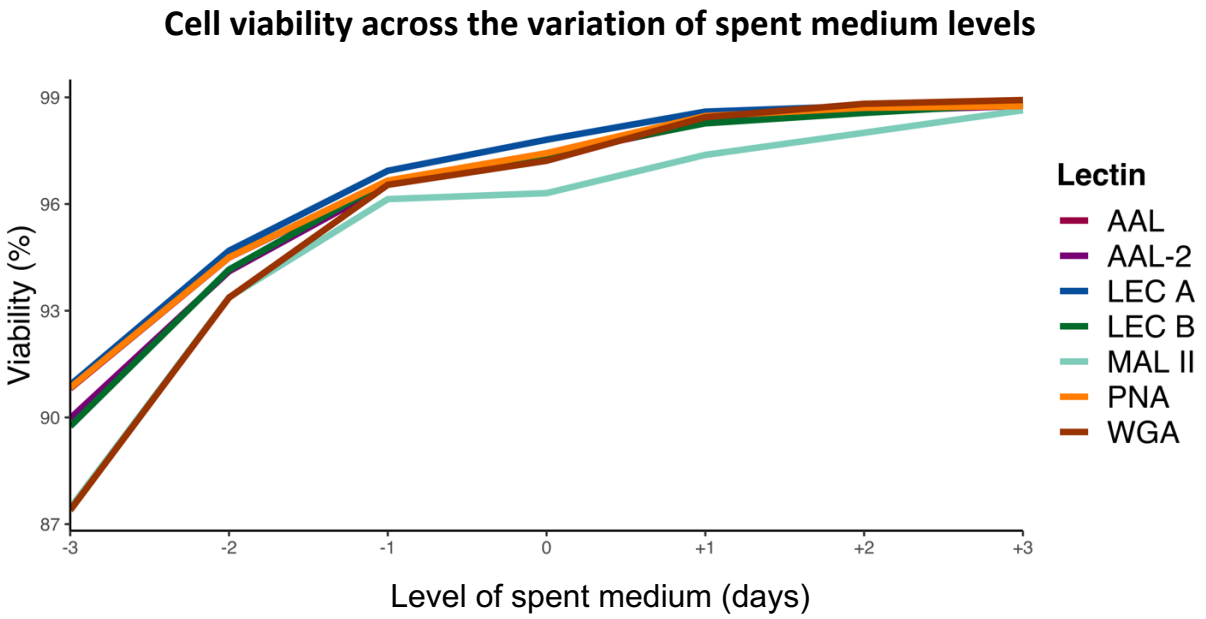

Figure 4.31: Line plot showing polynomial fits of cell viability of all lectin samples from the experiment involving the variation of spent medium levels. Data obtained through a BD FACSAria™ I flow cytometer and using 7-AAD to stain dead cells, thus determining cell viability (see section 3.12.1.6).

As illustrated in the figure, the tendency of a sharp decrease in cell viability for further depleted nutrient levels was experimentally confirmed since the vast majority of cells cultivated at -4, -5 days and so on, were dead. As a consequence, flow cytometric data from these cultures could not be obtained. In fact, even the -3 day cultures (triplicate) had most of the cells compromised. This is discussed in more detail when sample size and power analysis for the spent medium variation experiment is covered.

Figure 4.32 shows the change in pH as the level of spent medium was varied. It can be observed that the pH decreased up to -1 day point, but it stabilized at 7.3 between -1 and 0 day. Further decreases in spent medium levels resulted in pH values lower than 7 pH. The decrease rate between -3 and -1 days was higher than between 0 and 3 days, demonstrating a tendency to rapidly accumulate alkaline compounds in the medium for further increases in spent medium levels.

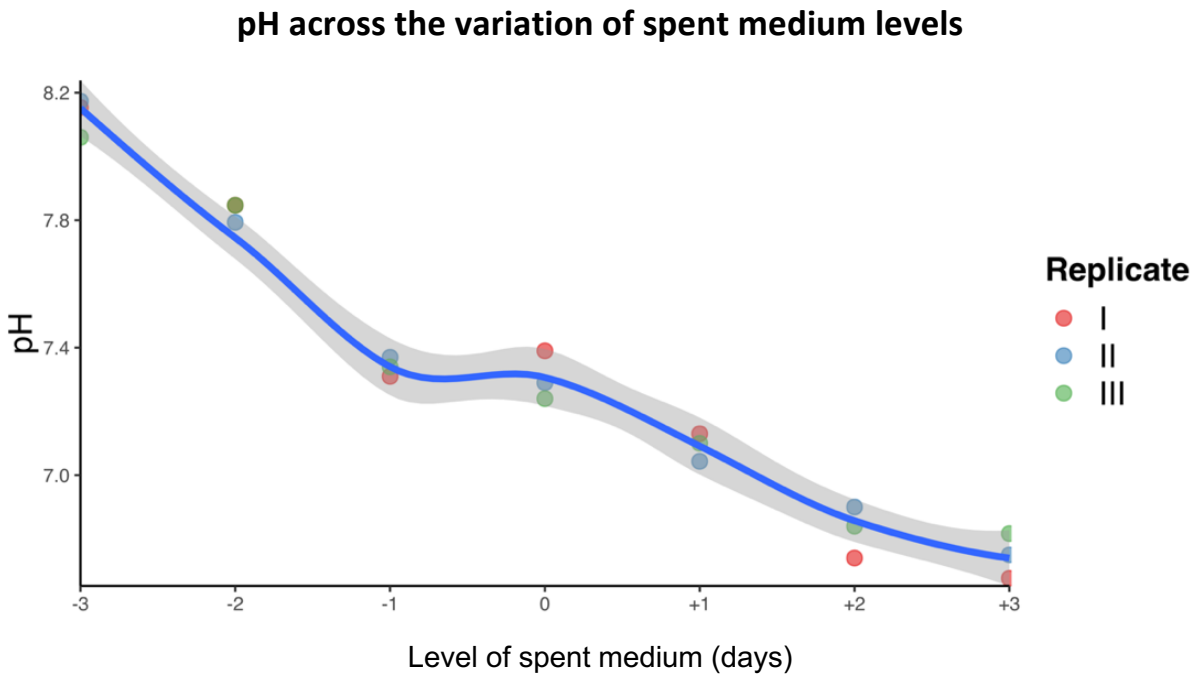

Figure 4.32: Line plot showing a polynomial fit with 95% confidence interval of pH as a function of the variation of spent medium levels. The values of pH were obtained using an Orion Semi-micro pH electrode (see section 3.12.1.8).

By observing both figures from -1 day to -3 days, it can be concluded that the cell viability decreases sharply due to the rapid increase in pH within this range. The pH value of 7.3 is an optimal one allowing the cell viability to remain above 96%. Whereas, pH higher than 7.4 causes a rapid reduction in cell viability due to the accumulation of alkaline metabolites released by the cells and glutamine degradation into ammonium (Borys, Linzer and Papoutsakis, 1994; Slivac *et al.*, 2010).

Descriptive analysis of the variation of the relative cell size and cell internal and external complexity parameters

Flow cytometry allows the extraction of information on the relative cell size (FSC channels) and the relative internal and external cell complexity (SSC channels) (see section 1.3.2). Therefore, a descriptive analysis of the FSC-A and SSC-A of the dead and apoptotic cell subpopulations as well as the DNA subpopulations is developed in this section. The plots are facetted by the 7 lectins and a common FSC-A and SSC-A scales are used to facilitate the direct comparison among the lectin panel. Additionally, lectin facets are organized in pairs of a commercial lectin and its counterpart in-house lectin (see section 4.7).

Figure 4.33 demonstrates the variation of the relative cell size mean as cells were subjected to a different level of spent medium in the last 24 hours of the cell culture process. The figure shows 5 subpopulation curves per lectin. For all lectins, it can be clearly seen that the pattern of the change in FSC-A mean was the same between the baseline and +3 days. However, two patterns emerged between -3 days and the baseline, particularly in relation to the DNA subpopulations.

The first pattern can be observed in AAL, PNA, LEC A and WGA curves in which the FSC-A means of the DNA cell cycle subpopulations increased (from -2 to 0 day) after a decrease between -3 and -2 days. Whereas, the second pattern observed in LEC B, AAL-2 and MAL II curves, a constant increase in the means was seen up to 0 day. Such pattern could be due to the fact that most of the cells from the -3 day-cultures were compromised owing to the low levels of nutrients in the medium, as has been previously pointed out (Figures 4.31 and 4.32). Overall, it can be observed a decrease in the relative cell size as the cells were subjected to increasing depleted nutrient levels. Also, the positions of the relative cell size curves in relation to FSC-A scale are slightly different across the lectins, demonstrating that the lectin interaction on the surface of the cell might not significantly influence the FSC-A signal or all lectins from the panel influence the signal at the same degree.

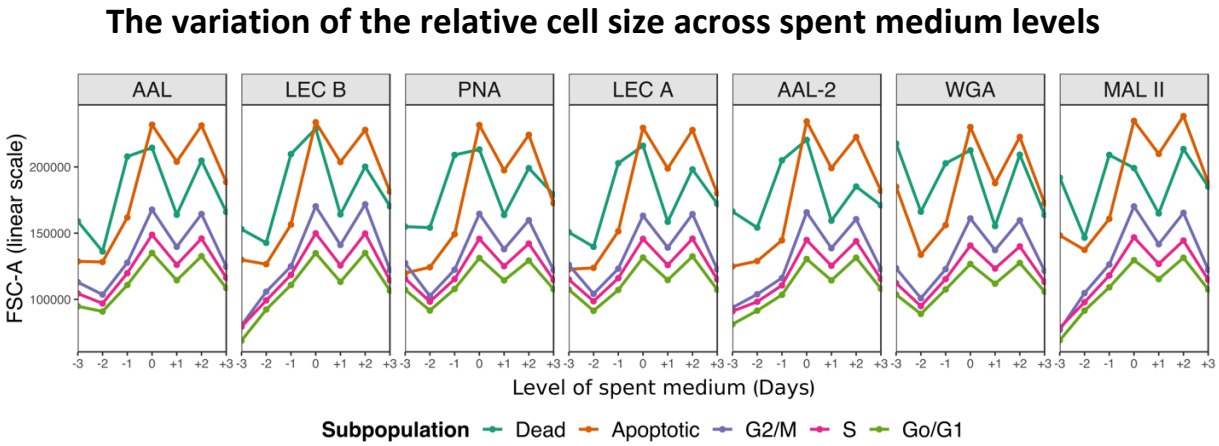

Figure 4.33: Lectin-facetted line plot showing the relative cell size (FSC-A) variation as a function of the spent medium level for 5 different cell subpopulations: dead, apoptotic, and DNA cell cycle subpopulation (G2/M, S, and Go/G1). Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.8).

In addition, the figure allows the observation of the increasing cell size as the cells go through the DNA cell cycle. This can be observed by the position of the curves within each lectin facet. Go/G1 < S < G2/M < Apoptotic < Dead is the ascending cell size order which can be observed in all lectin facets at depleted (negative) nutrient levels. However, dead cell size decreased significantly at positive nutrient levels and the size order changed to Go/G1 < S < G2/M < Dead < Apoptotic. Therefore, it can be observed that as cells went through the DNA cycle their relative cell size increased. For instance, during the DNA cell cycle, cells are in the process of duplicating the DNA (Go/G1 and S) reaching two sets of genetic material (G2/M) when mitosis takes place. Thus, this process alone changes the size of a cell (see section 1.4).

Figure 4.34 shows the alteration of the relative cell internal and external complexity as the level of spent medium was varied. The upper plot shows 5 subpopulation curves for each lectin, while the lower plot highlights the curves of the DNA subpopulations, facilitating the observation of the trends revealed by these curves. Overall, all lectin curves showed a very similar pattern except for LEC B, AAL-2 and MAL II that showed a different pattern at -3 day spent medium level, as was also observed in the data of FSC-A signals (Figure 4.33).

It can be observed from the upper plot that the order of SSC-A signal magnitude of the subpopulation curves for the lectin panel is Go/G1 < S < G2/M < Apoptotic < Dead as was also observed in the figure 4.33 from -3 to 0 day levels. However, between 0 and +1 day, the dead and apoptotic curves showed a tendency to equalize their SSC-A signals. MAL II curves in particular, showed that the apoptotic curve surpassed the dead curve between 0 and +1. Therefore, it can be concluded that as cells go through the DNA cycle their relative internal and external complexity level increases. This conclusion is expected as the extra genetic material being synthesized by the cell during the DNA cell cycle increases the number of biomolecules within the cell, thereby elevating the complexity of the internal environment (Ozlu *et al.*, 2015; Ly *et al.*, 2017).

All lectin subpopulation curves showed that the SSC-A signal fluctuated around the same values at the positive levels of spent medium. However, while dead and apoptotic curves showed a considerable decrease in SSC-A at negative spent medium levels (depleted nutrient levels), Go/G1, S and G2/M SSC-A signals increased at the first spent medium levels, but a tendency to decrease the signal was observed as cells were subjected to further negative levels of spent medium. In other words, cells became internally and externally more complex, demonstrating possibly an increase in cellular metabolism to cope with the first effects of nutrient depletion up to a depletion level which was low enough to lower down cellular metabolism and eventually lead cells to death. Although the extra levels of nutrients did not lead cells to death, excess of nutrient might also have caused metabolic changes in relation to the baseline nutrient level.

The variation of the relative cell internal and external complexity parameters spent medium levels

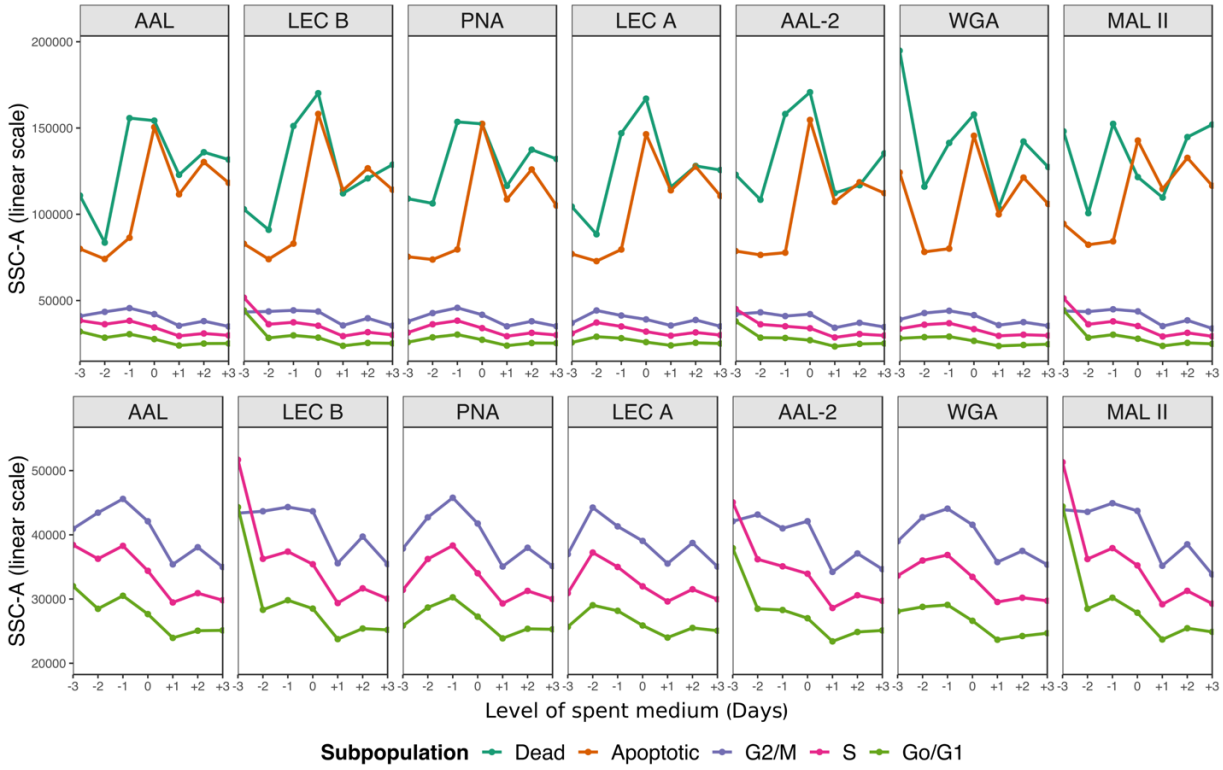

Figure 4.34: Lectin-facetted line plots demonstrating the relative cell internal and external complexity (SSC-A) variation as a function of the spent medium level for 5 different cell subpopulations: dead, apoptotic, and DNA subpopulations (G2/M, S, and Go/G1). The lower plot highlights curves of the DNA subpopulations to better visualize their trends. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.8).

It has been shown that both nutrient deprivation and nutrient excess can cause cellular stress (Wellen and Thompson, 2010). The uptake of nutrients in mammalian cells is controlled primarily by growth factor signaling which is associated with the level of reactive oxygen species produced by mitochondria. Reactive oxygen species are produced at a low level, allowing the normal cellular functioning. However, these species can rise with alterations in oxidative mitochondrial metabolism due to both deprivation and excess of nutrients for instance. This rise can potentially cause damage to the components of a cell and its death

(Veal, Day and Morgan, 2007; Trachootham, Alexandre and Huang, 2009; Hamanaka and Chandel, 2010). In the case of cellular stress caused by the excess of nutrients, in diseases characterized by alterations in cellular metabolism such as cancer and diabetes, increased levels of reactive oxygen species are found (Wallace, 2005; Brandon, Baldi and Wallace, 2006; Halliwell, 2007; Nathan, 2008; Roberts and Sindhu, 2009; Trachootham, Alexandre and Huang, 2009).

Statistical analysis of the variation of cell surface glycoprofile

The statistical analysis of the variation of cell surface glycoprofile across different levels of spent medium is composed of three stages: descriptive, inferential and power analysis.

This section firstly investigates the patterns and tendencies revealed by the variation of the means of LECTIN-A detector channel signal as the levels of spent medium were altered. Line plots are shown with a common LECTIN-A scale to facilitate the comparison of lectin interaction intensities across the lectin panel. Secondly, the analysis of the levels of statistical significance of the difference between each treatment (+1 day, +2 days, -1 day, -2 days for instance) and the baseline dataset is developed. In addition, this analysis allows the observation of data variability since box plots are demonstrated. Lastly, in order to evaluate whether the findings are scientifically meaningful, the results of power analysis are discussed.

Figure 4.35 demonstrates the variation of LECTIN-A signal as the levels of spent medium changed. Consequently, the variation of cell surface glycoprofile can be investigated. As was observed previously on the data from FSC-A and SSC-A detector channels, the order of the LECTIN-A signal intensity of the subpopulations of all lectins is: Go/G1 < S < G2/M < Apoptotic < Dead. It can be concluded that as cells go through the DNA cell cycle, the quantity of lectin binding sites increases possibly due to cell size enlargement as was observed from the Figure 4.33. This conclusion can be further supported by the fact that G2/M and S subpopulation curves of each lectin revealed a repetition of the Go/G1 curve pattern.

The increasing lectin binding detected from apoptotic and dead cells was expected, since apoptotic cells gradually lose the integrity of the membrane, compromising the osmotic regulation of the cell. Eventually, cells are fully dead allowing the absorption of molecules in high concentration levels outside of the cell.

The bottom plot of the figure allows the identification of the strongest and weakest LECTIN-A signal from the DNA subpopulations. WGA was demonstrated to be the strongest binding on the cell surface and LEC A the weakest. Therefore, based on the studies shown in section 4.7, it can be concluded that the quantity of N-Acetylglucosamine groups available for binding is at a higher number in relation to Fucose, Sialic acid and Galactose, particularly in relation to Galactose. This was demonstrated by the lowest signal detected which was from LEC A followed by PNA, both of which interact with Galactose.

Although AAL-2 also binds to N-Acetylglucosamine, the lectin did not show the same efficiency at binding to the sugar molecule as WGA did. However, this could be due to the difference in the amount of biotin molecules each lectin has. Since those lectins have gone through different biotinylation processes, as one was a commercial molecule and the other was an in-house molecule. Nonetheless, although the intensities can be different, the pattern of the curves of an in-house lectin and its commercial equivalent were generally very similar, demonstrating that the lectins were interacting with the same sites on cell surface. This observation further supports the results obtained in section 4.7.

The variation of cell surface glycoprofile across spent medium levels

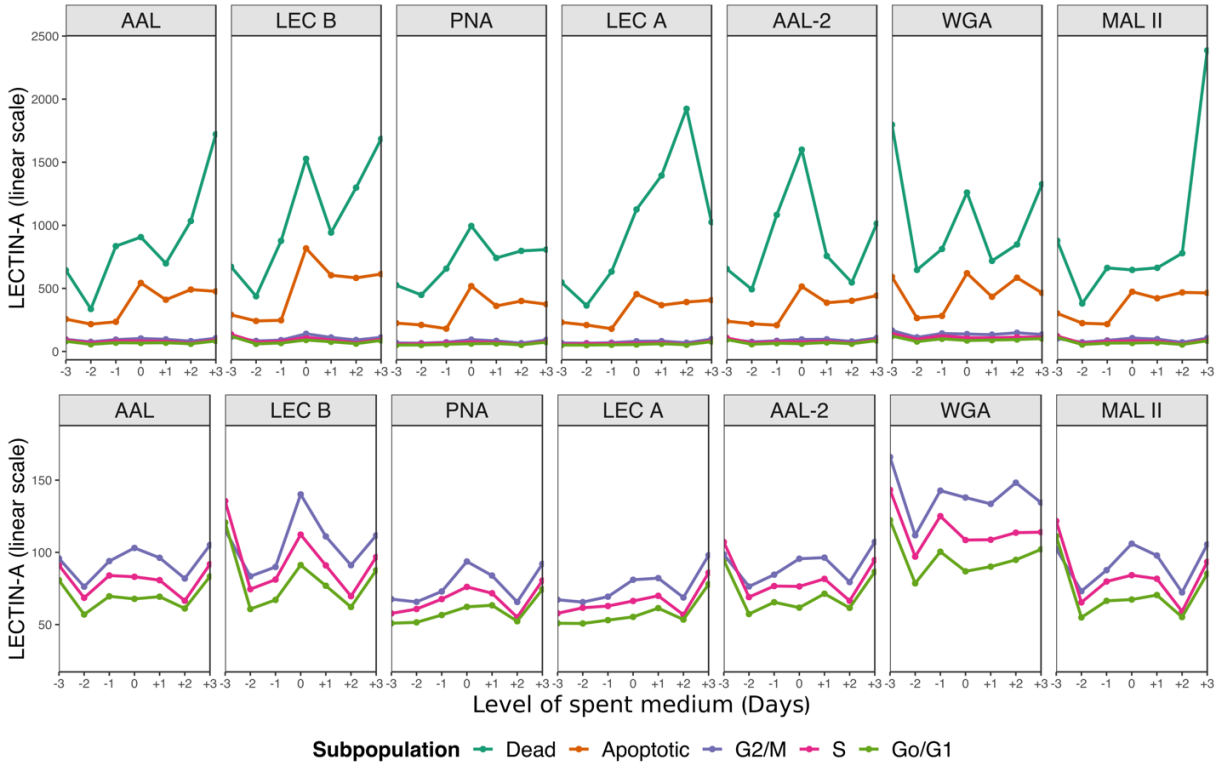

Figure 4.35: Lectin-facetted line plots demonstrating the lectin interaction (LECTIN-A) variation for 5 different cell subpopulations: dead, apoptotic, and DNA subpopulations (G2/M, S, and Go/G1). The lower plot highlights the curves of the DNA subpopulations to better visualize their trends. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.8).

With the purpose of evaluating the statistical significance of the glycoprofile differences observed in relation to the baseline nutrient level, inferential analysis was conducted and the data was plotted in the form of a box plot colored with the level of the statistical significance as shown in Figure 4.36 (see section 3.12.7).

The figure shows the data distributions of the treatments applied (levels of spent medium), including the baseline. However, the baseline boxplot is colored in grey to highlight it as the data distribution to which the different treatments were compared (see section 3.12.7). The investigation of the variation of cell surface glycoprofile of live cells is the main goal of this

research work; thus, the DNA subpopulations are the cells of most interest. Consequently, the figure shows data of G2/M, S, and Go/G1 subpopulations.

Inferential analysis of the cell surface glycoprofile variation across different levels of spent medium

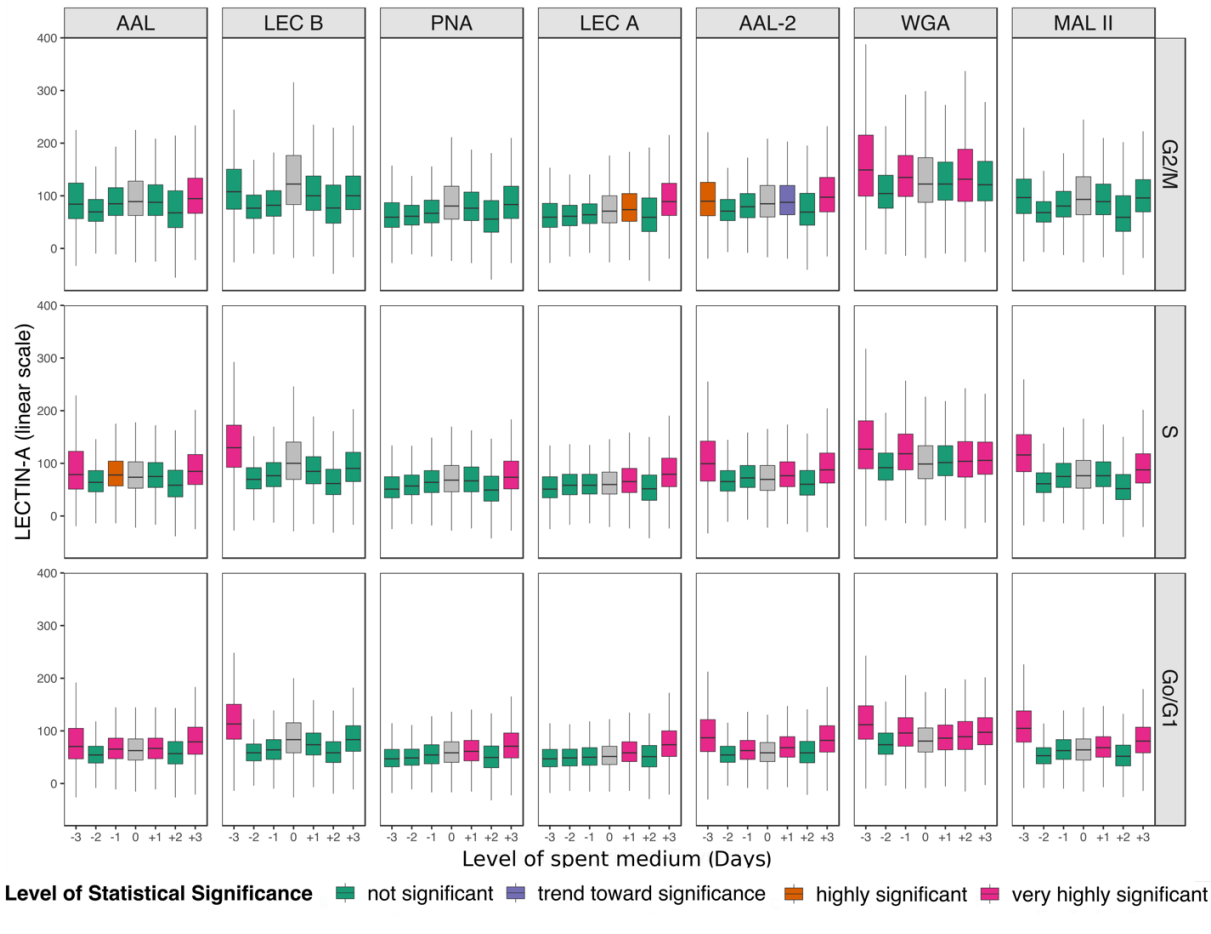

Figure 4.36: Box plot facetted by lectin and DNA cell cycle subpopulations highlighting the levels of statistical significance of the glycoprofile difference between the treatments and the baseline. Data obtained through a BD FACSAria™ I flow cytometer using LECTIN-A detector channel (see sections 3.12.1.8 and 3.12.7).

Figure 4.36 allows the observation of the highest number of *very highly significant* changes detected in the Go/G1 subpopulation. Additionally, the figure shows a reduction in the

number and in the level of statistical significance of the S subpopulation in relation to Go/G1. Likewise, this can be seen in the comparison between G2/M and S subpopulations. For instance, at +1 day level, G2/M cells of AAL-2 demonstrated a *trend toward significance*, while the S and Go/G1 cells interacting with the same lectin demonstrated *very highly significant* changes in the cell surface glycoprofile in relation to the baseline glycoprofile. This was also observed at +1 day level of LEC A subpopulations and at -3 day level of AAL-2 subpopulations. The sample size difference can be the statistical reason why the highest number of *very highly significant* changes was computed in Go/G1; thus such information is evaluated later in this section. Therefore, the further analysis is focused on Go/G1 subpopulation.

The Go/G1 subpopulation demonstrated that WGA detected the highest number of *very highly significant* changes at 5 treatments out of the 6 applied to the cells. AAL-2, the in-house WGA counterpart, demonstrated the second highest number, 4 treatments, just the same as AAL. LEC B showed the lowest number of *very highly significant* change detected for only one treatment, -3 day nutrient level. PNA and LEC A detected 2 and MAL II detected 3. Table 4.4 summarises the number of *very highly significant* changes for each lectin in the Go/G1 subpopulation. Most of the *very highly significant* changes was found in the treatments which cells were treated with extra levels of nutrients. However, cells subjected to depleted nutrient levels showed some dramatic differences in the glycoprofile, particularly at -3 day treatment (Figure 4.35).

Table 4.4: Table summarizing the number of *very highly significant* changes detected by each lectin and the nutrient treatments in which these changes were found in the Go/G1 subpopulation.

Lectin	Number of *very highly significant* changes	Treatments in which significant changes were found
WGA	5	-3, -1, +1, +2, and +3 days
AAL-2	4	-3, -1, +1, and +3 days
AAL	4	-3, -1, +1, and +3 days
MAL II	3	-3, +1, and +3 days
PNA	2	+1 and +3 days
LEC A	2	+1 and +3 days
LEC B	1	-3 days

In order to assess whether the glycoprofile changes are scientifically meaningful, power analysis was performed. Figure 4.37 shows the results of the analysis covering the DNA subpopulations. It can be observed that the highest power values are in the data obtained from Go/G1 cells. This observation is in agreement with what was observed in Figure 4.36, which showed the highest number of *very highly significant* changes in Go/G1 cells.

Figure 4.37 reveals that some of the glycoprofile changes in the Go/G1 subpopulation are quite scientifically meaningful both in the treatments in which cells were depleted of nutrients and with a nutrient excess in relation to the baseline nutrient level. This can be observed in the data of MAL II, LEC B and PNA for instance. In the case of LEC B, computed power values for -2 and +2 day treatments were higher than 75%; the former power value was nearly 100%. In other words, if an experiment with 75% power was to be repeated 100 times, the methodology would be able to find a statistically significant change when there is one, in 75 times.

Power analysis of the glycoprofile changes detected across spent medium levels

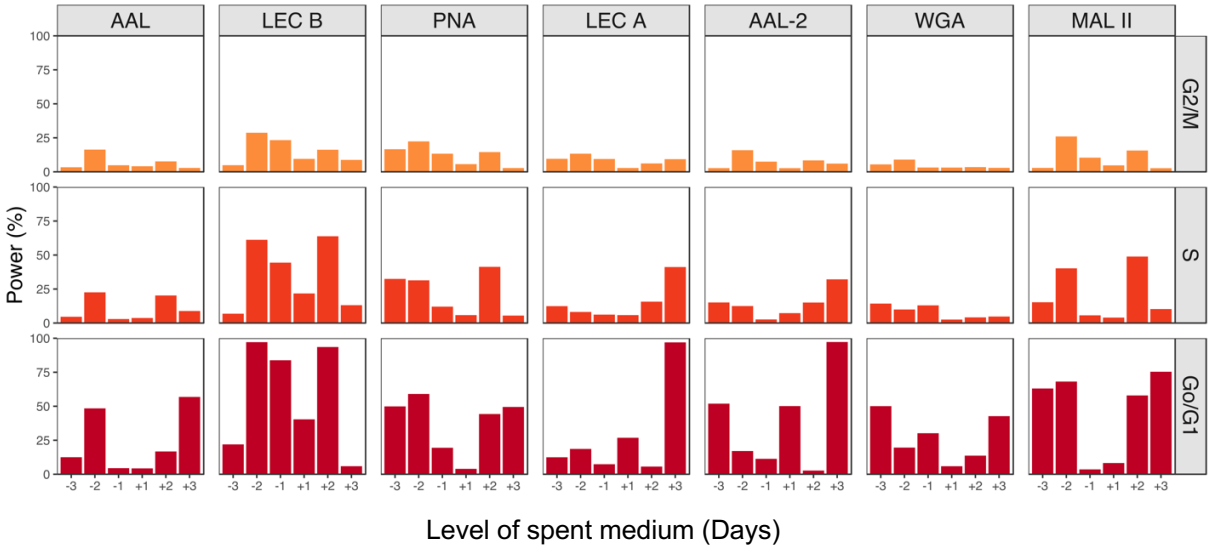

Figure 4.37: Bar plot facetted by lectin and DNA subpopulations demonstrating the results of power analysis of the cell surface glycoprofile differences which were detected between the multiple levels of spent medium (treatments) and the baseline. Powers analysis on the data obtained through a BD FACSAria™ I flow cytometer (see sections 3.12.1.8 and 3.12.7).

The power is influenced by the sample size as was outlined in section 3.12.7.3. Thus, Figure 4.38 demonstrates the number of cells used in each treatment to perform the statistical analysis. As previously pointed out, Go/G1 cells were in the highest number, thereby increasing the power computed from this subpopulation. The figure also shows a decline in the sample size of treatments in which cells were subjected to depleted nutrient levels. Such reduction in sample size is a reflection of the sharp decrease in viability shown in Figure 4.31.

Although the reduction in sample size decreases the power, a considerable glycoprofile difference causes an increase in the power value (see section 3.12.7.3). This can be particularly observed in the power value of MAL II at -3 and -2 days treatments. Data variability is another important factor influencing the power: if the variability increases, then the power decreases.

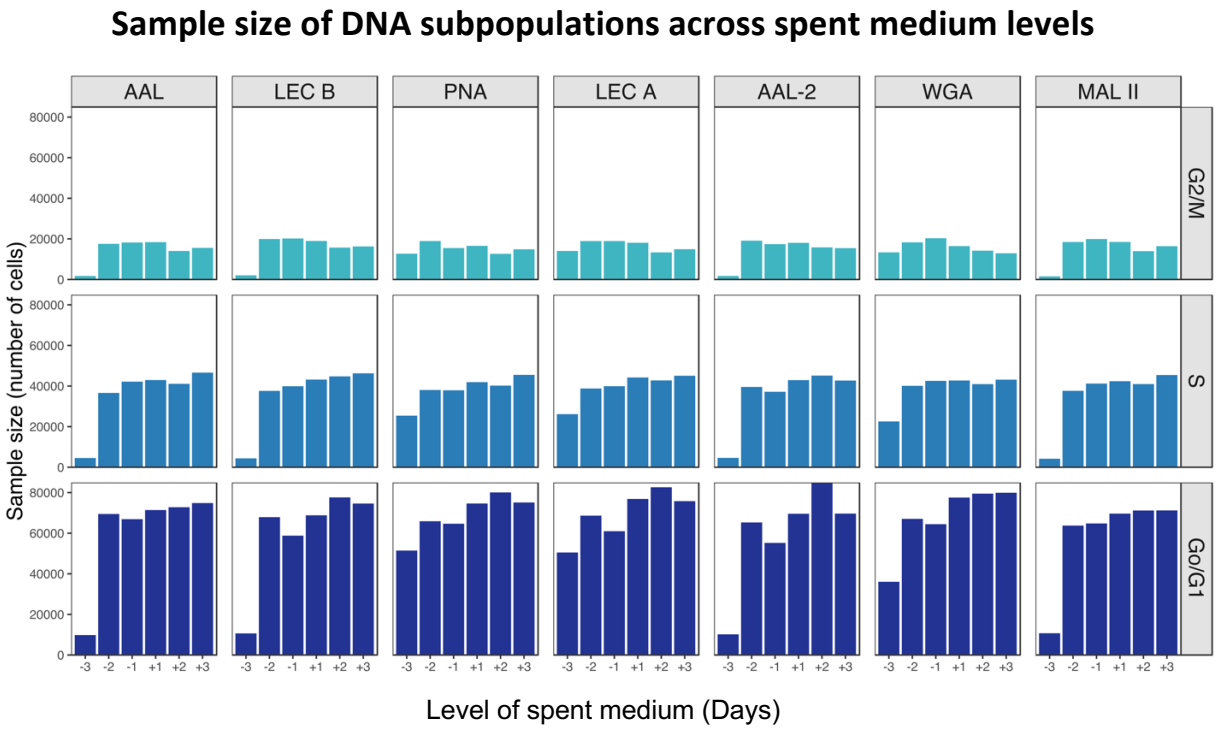

Figure 4.38: Bar plot facetted by lectin and DNA cell cycle subpopulation demonstrating the sample sizes used in the statistical analysis of cell cultures treated with different levels of spent medium. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.8).

Since the sample sizes were generally the same among the nutrient treatments (except for -3 day treatment), the power results were mostly influenced by the difference between a treatment and the baseline data variabilities as well as by the difference in the means of the treatment and baseline. With the purpose of discussing the relationship of power, data variability difference and the difference in the means, a more detailed figure was generated containing a bar plot with treatment power values and a box plot overlaid with the curve constructed with the means of LECTIN-A signals of each treatment. The plots can be observed in Figure 4.39.

Although the inferential analysis has shown more *very highly significant* changes in cells treated with extra levels of nutrients (+1, +2, and +3) as demonstrated by the lower plot, power analysis (upper plot) has revealed that both depleted and extra levels of nutrients, in fact, can equally cause significant impacts on cell surface glycoprofile. For instance, LEC B

power results for depleted and extra levels of nutrients are very scientifically meaningful with 97% at -2 days and 93% at +2 days of nutrient levels in relation to baseline. Although with lower power values, MAL II is another example in which both depleted and extra nutrient levels impacted the glycoprofile, with 63 and 68% power at -3 and -2 days respectively, and 58 and 75% at +2 and +3 days respectively.

Statistical analysis of Go/G1 cell surface glycoprofile across spent medium levels

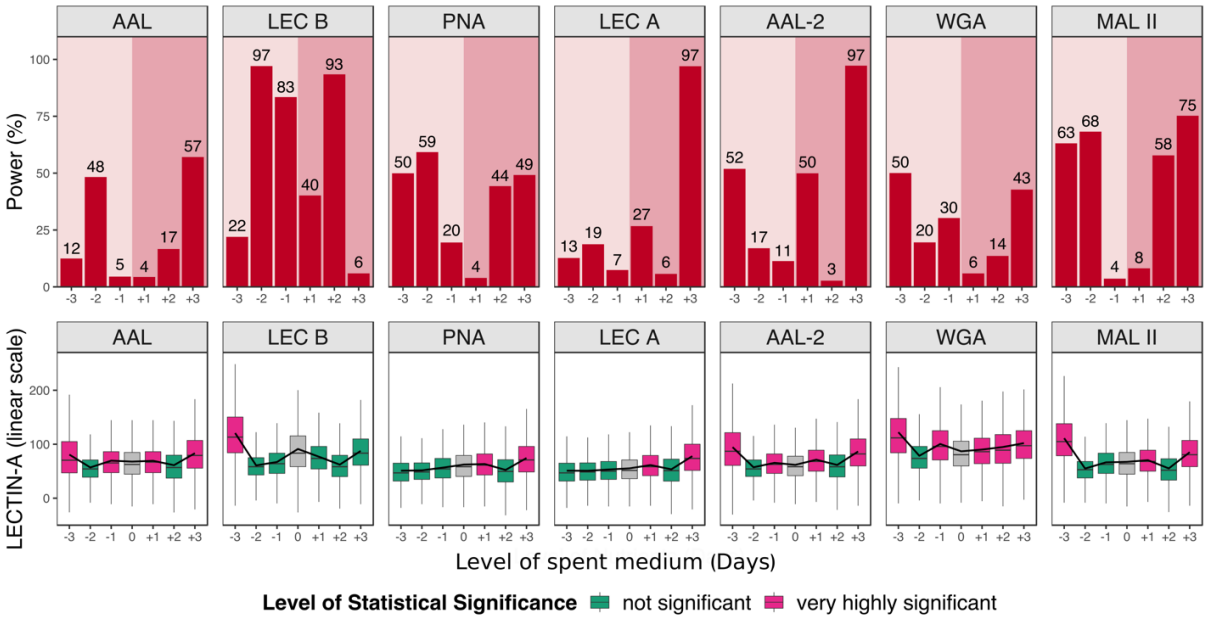

Figure 4.39: Complementary plots summarizing descriptive, inferential and power analysis of spent medium treatments applied to Go/G1 cells. **Upper plot:** bar plot facetted by lectin showing power values of each spent medium level and highlighting the positive and negative levels using different shades of red. **Lower plot:** Box plot color coded with the results of the inferential analysis and superimposed with a curve constructed by connecting the means of LECTIN-A signal of each spent medium level. Data obtained through a BD FACSAria™ I flow cytometer (see sections 3.12.1.8 and 3.12.7).

LEC A and AAL-2 clearly showed that the nutrient excess can also cause dramatic changes on cell surface glycoprofile. For instance, the glycoprofile changes detected by these lectins bound to the surface of cells treated with extra nutrient levels are more scientifically meaningful than the changes detected on cells subjected to depleted levels of nutrients. This was particularly demonstrated by LEC A in which the highest power value from a depleted nutrient level was 19% (-2 days), while 97% is the highest value from an extra nutrient level (+3 days). The lower plot shows no considerable changes in LEC A data variabilities and a fairly straight curve in the depleted levels, demonstrating that the low levels in power are due to the unchanged number of Galactose sites available for binding even though cells were subjected to increasing depleted nutrient levels. On the other hand, extra levels of nutrients increased the power values and the lower plot shows this increase is due to mostly a rise in the means even though the data variabilities have also increased. Therefore, it can be concluded that by increasing the nutrient levels, the number of available Galactose sites on the cell surface also increases, while depleted levels do not significantly influence the number of Galactose sites.

It can be generally concluded that by increasing the levels of nutrients up to 3 days, the number of available Fucose sites on cell surface increases and this increase is very scientifically meaningful with a power value that can go up to 94% detected with the aid of LEC B and up to 57% detected with AAL. The number of available sites of Galactose, N-Acetylglucosamine and Sialic also increases as a result of extra levels in the medium. With the aid of PNA, changes in the number of Galactose sites were detected with up to 49% and up to 97% with the aid of LEC A. N-Acetylglucosamine increase was detected by WGA with up to 43% power, while AAL-2 with up to 97%. Lastly, MAL II detected the increase in the number of available Sialic sites with up to 75% power. As a result of those high power values, it can be concluded that the increase in the number of available sites of Fucose, Galactose, N-Acetylglucosamine, and Sialic acid on the cell surface due to the rise in the levels of nutrients in the medium is scientifically meaningful.

On the other hand, as cells are subjected to depleted nutrient levels, available Fucose sites increase. However, LEC B was more powerful than AAL at detecting this increase with up to

97% power, while AAL detected it with up to 48%. Sialic acid sites also increase with up to 68% power detected with the aid of MAL II. Galactose and N-Acetylglucosamine sites have shown a tendency to increase; however, the power values were not as scientifically meaningful as the values obtained from the changes detected in the number of Fucose and Sialic acid. The highest power for Galactose was 59% by PNA and 19% by LEC A, and for N-Acetylglucosamine power was 52% by AAL-2 and 50% by WGA. Therefore, it can be concluded that the increase in the number of Fucose and Sialic acid as a result of depleted nutrient levels up to -3 days is scientifically meaningful. However, the changes in the number of Galactose and N-Acetylglucosamine are scientifically meaningful to a lower extent.

A number of scientific studies has demonstrated the association of glycosylation with nutrient levels. For instance, protein *N*-linked glycosylation and *O*-linked N-Acetylglucosamine (*O*-GlcNAc) alteration can be regulated by nutrient levels through the flux of glucose into the hexosamine biosynthetic pathway which is highly in tune with cellular metabolism. The hexosamine pathway end product is UDP-GlcNAc, which is essential for both nucleocytoplasmic *O*-GlcNAc protein modification and *N*-linked glycosylation in the Endoplasmic reticulum and Golgi (Love and Hanover, 2005; Dennis, Nabi and Demetriou, 2009b). Galactosylation and sialylation levels of camelid-humanized monoclonal antibody expressed in CHO cells increases due to lower levels of glutamine in comparison to the protein expressed in higher levels of glutamine (Aghamohseni *et al.*, 2014).

Since cell viability was compromised and medium pH rose in response to the depleted levels of nutrients, it can be concluded that the increase in the number of available Fucose, N-Acetylglucosamine and Sialic acid sites and the decrease in Galactose sites on CHO-K1 cell surface can be an indicative of changes in the metabolism of the cell which are associated with the lack of nutrients in the medium. Therefore, an increase in the cell internal and external complexity level and a decrease in the relative cell size are also alterations associated with the suboptimal levels of nutrients.

4.9.2 Analysis of the effects of temperature variation

The experimental data involving the variation of temperature levels comprises 9 datasets each representing a temperature level: 8 datasets consist of the variation of the temperature levels in the last 24 hours of cell culture, and 1 dataset extracted from cells grown under the temperature baseline condition throughout the entire 96 hours of culture (sample collection point).

In order to facilitate the understanding of the data, the variables of interest were plotted against the variation of temperature in relation to the setpoint or the baseline condition. In other words, the 0 point of the x axis (Temperature) depicts the data obtained from cells cultivated at temperature baseline condition, whereas the remaining points depict the datasets obtained from the different levels of temperature measured in one unit of degree Celsius. The experimental setup is fully described in section 3.12.2.3.

Cell culture parameters

It can be observed through Figure 4.40 that cell viability for all lectin samples remained stable within 90 to 94% up to baseline point. Thus, the reduction in temperature up to below 5 units (32°C) from the set point did not influence the cell viability. However, a significant decrease in cell viability is observed when cells were treated with increasing temperature levels reducing cell viability to values below 70%, revealing the harmful effect of high temperature levels on the cell culture process. The impact of this dramatic reduction in cell viability is later discussed when power analysis and sample size are looked at in more detail.

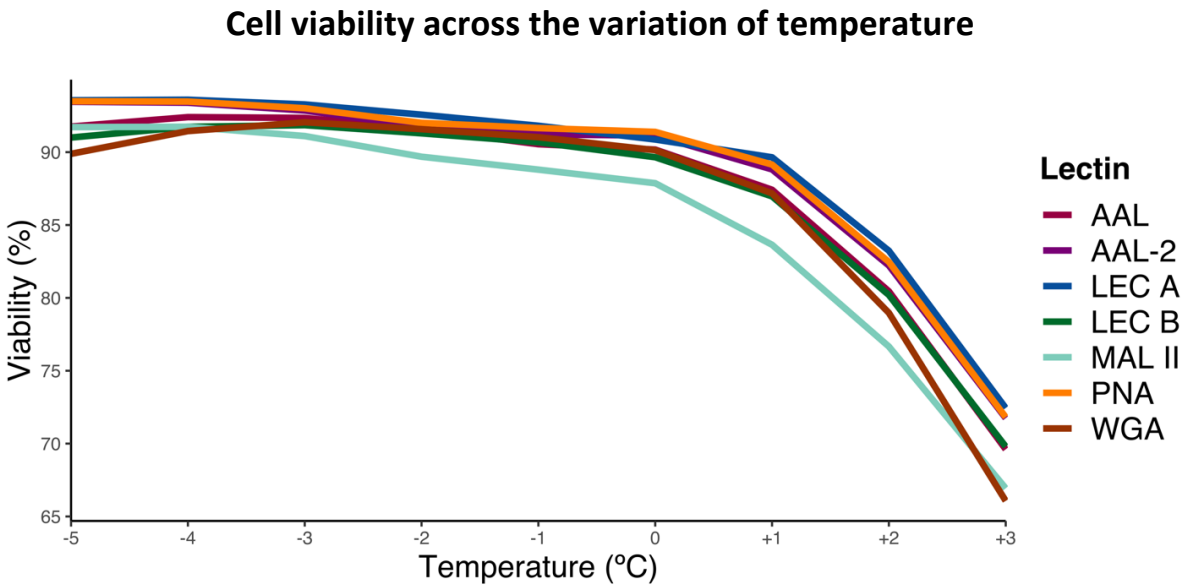

Figure 4.40: Line plot showing polynomial fits of cell viability of all lectin samples from the experiment involving the variation of temperature. Data obtained through a BD FACSAria™ I flow cytometer and using 7-AAD to stain dead cells, thus determining cell viability (see section 3.12.1.6).

Figure 4.41 shows the change in pH as temperature was varied. It can be observed that pH values were within 7.25 to 7.50 up to baseline point. Between 0 and +1°C, the pH dropped below 7.25 and subsequently increased sharply as temperature increased, reaching 7.75 pH at +3°C. Thus, likewise the cell viability, the pH of the cell cultures subjected to up 5°C units below the baseline condition was not affected. However, temperatures up to 3°C units above 37°C caused a decrease followed by a sharp increase in pH.

Decreased temperatures in animal cells have been demonstrated to reduce cellular metabolism, glucose and glutamine consumption, free radical oxygen species, to inhibit the release of metabolic waste, to arrest the cell cycle mainly at G1 phase, to increase cell viability and delay apoptosis (van Breukelen and L. Martin, 2002). Many of the cellular responses to decreased temperatures are also observed in responses to increased temperatures (Al-Fageeh and Smales, 2006). However, elevated temperatures in medium supplemented with L-glutamine accelerate its degradation rate, increasing the concentration of ammonia and

therefore raising the pH (Borys, Linzer and Papoutsakis, 1994; Slivac *et al.*, 2010). Therefore, the sharp increase in pH between +1 and +3 might be mostly due to the accumulation of alkaline compounds derived from the L-glutamine accelerated degradation.

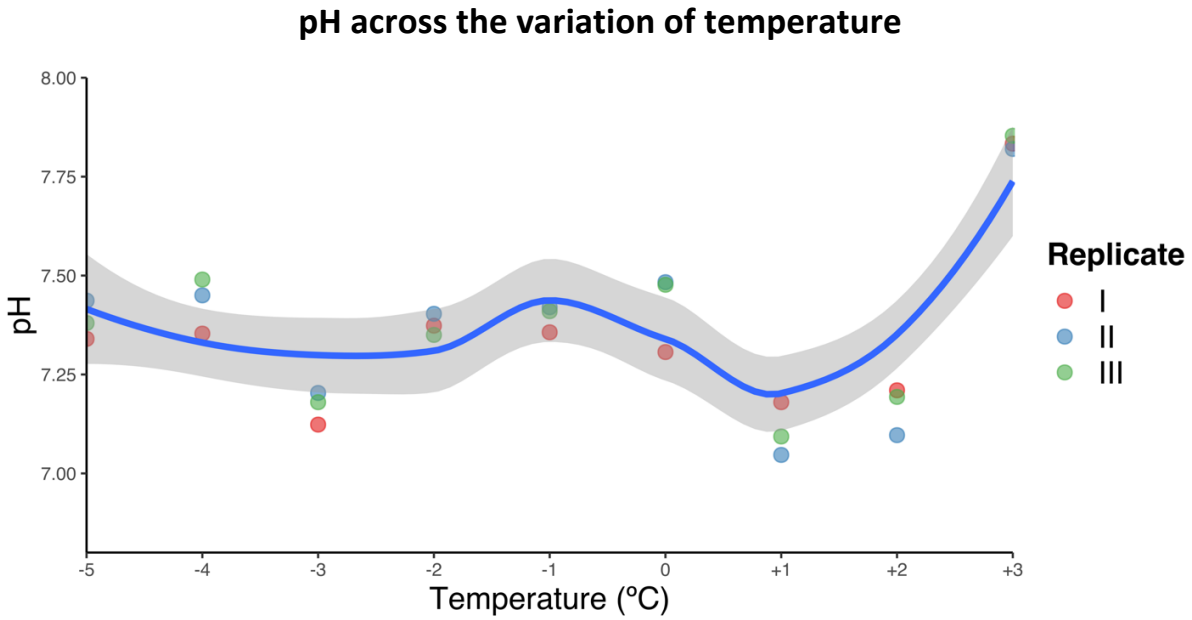

Figure 4.41: Line plot showing a polynomial fit with 95% confidence interval of pH as a function of the variation of temperature. The values of pH were obtained using an Orion Semi-micro pH electrode (see section 3.12.1.8).

To conclude, the observation of both figures allows the identification of +1 to +3°C as the temperature variation interval in which cell viability rapidly diminishes alongside a dramatic increase in pH. Therefore, the rise in pH, as was also observed in the previous section (Figure 4.32) may be strongly associated with the decrease in cell viability.

Descriptive analysis of the variation of the relative cell size and cell internal and external complexity parameters

Figure 4.42 demonstrates the variation of the relative cell size mean as cells were subjected to a different level of temperature in the last 24 hours of the cell culture process. For all lectins, it can be clearly seen that the pattern of the change in FSC-A mean was the same, particularly in relation to Go/G1, S, G2/M and apoptotic curves. However, the DNA curves of PNA, LEC A, and AAL-2 showed a rapid increase in the relative cell size mean between +2 and +3°C.

Additionally, it can be observed that increased temperature levels did not substantially change the relative cell size of live cells, except for PNA, LEC A and AAL-2 subpopulations for which an increase in the FSC-A signal was detected between +2 and +3°C. On the other hand, reduced temperature levels caused a considerable increase in the relative cell size between 0 and -1°C interval followed by a decrease reaching the FSC-A signal levels detected at baseline point.

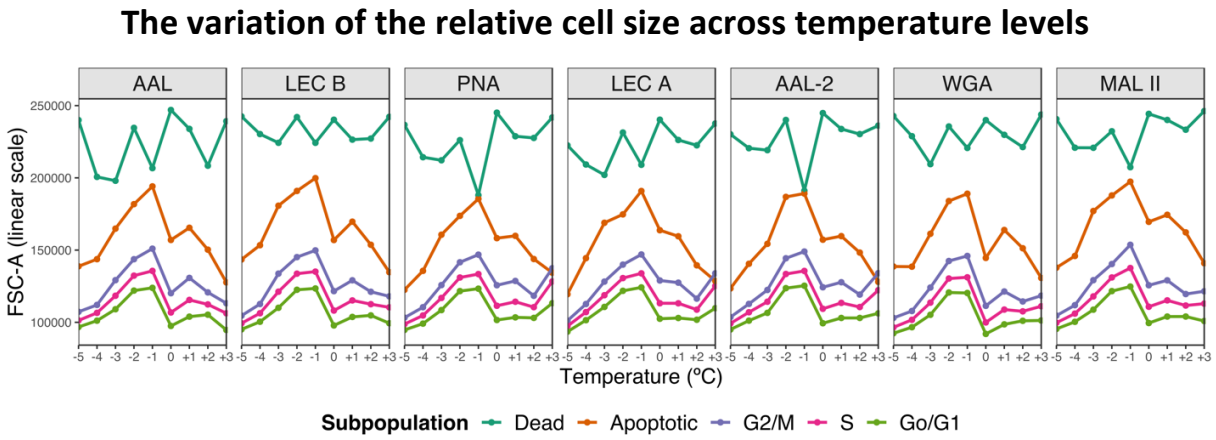

Figure 4.42: Lectin-facetted line plot showing the relative cell size (FSC-A) variation as a function of temperature for 5 different cell subpopulations: dead, apoptotic, and DNA subpopulations (G2/M, S, and Go/G1). Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.8).

Likewise it was previously observed from FSC-A data of the nutrient variation experiment (Figure 4.33), the positions of the relative cell size curves in Figure 4.42 in relation to FSC-A scale were slightly different across the lectins, demonstrating that the lectin interaction on the surface of the cell might not significantly influence the FSC-A signal or all lectins from the panel influence the signal to the same degree. Another similarity observed between the nutrient variation FSC-A and temperature variation FSC-A datasets is the cell size order across the subpopulations: Go/G1 < S < G2/M < Apoptotic < Dead.

Figure 4.43 shows the alteration of the relative cell internal and external complexity as temperature was varied. The upper plot shows 5 subpopulation curves for each lectin, while the lower plot highlights the curves of the DNA subpopulations facilitating the observation of the trends revealed by these curves. Overall, all lectin curves showed a very similar pattern, particularly the DNA curves which are the subpopulations of most interest since they were composed of live cells. A considerable change in SSC-A signal was not observed in temperatures below the baseline temperature level. However, considerable fluctuations were seen in temperatures above the baseline level and the most dramatic fluctuation was between +2 and +3°C in which a sharp increase in SSC-A mean signal was detected across the lectin panel. In other words, the relative cell internal and external complexity level increased substantially within this interval.

The variation of the relative cell internal and external complexity parameters across temperature levels

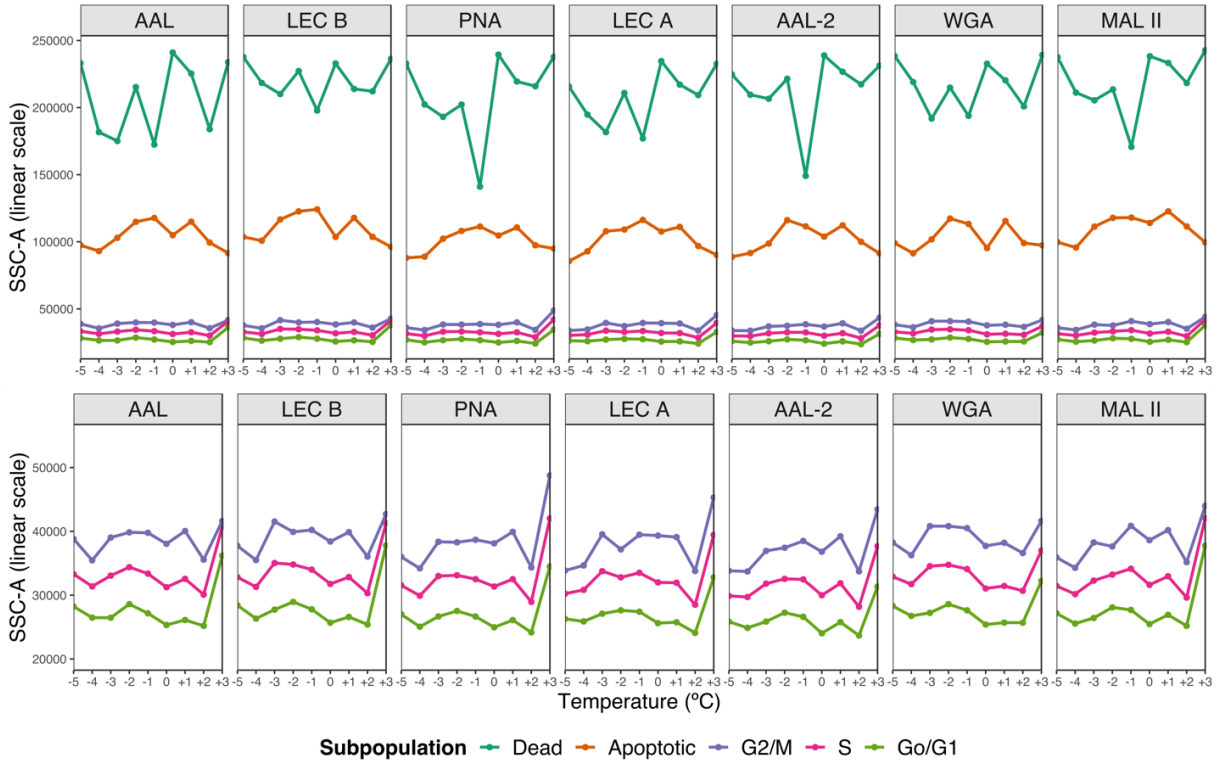

Figure 4.43: Lectin-facetted line plots demonstrating the relative cell internal and external complexity (SSC-A) variation as a function of temperature for 5 different cell subpopulations: dead, apoptotic, and DNA subpopulation (G2/M, S, and Go/G1). The lower plot highlights the DNA subpopulation curves to better visualize their trends. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.8).

Likewise Figure 4.34, 4.43 allows the observation of the same order of SSC-A signal magnitude of the subpopulation curves: Go/G1 < S < G2/M < Apoptotic < Dead. This further supports the previous conclusion on the increasing cellular metabolism level as cells go through the DNA cell cycle.

Statistical analysis of cell surface glycoprofile variation

The workflow of this section is the same adopted previously for the statistical analysis of the variation of cell surface glycoprofile across different levels of nutrients. Therefore, the investigative analysis of the cell surface glycoprofile variation due to temperature alteration initiates with a descriptive analysis of data obtained from LECTIN-A detector channel, then the inferential and power analysis of the data is developed.

Figure 4.44 demonstrates the variation of LECTIN-A signal as temperature changed. Therefore, the variation of cell surface glycoprofile can be investigated. As was observed previously on data from FSC-A and SSC-A detector channels, the order of the LECTIN-A signal intensity of the subpopulations of all lectins is: Go/G1 < S < G2/M < Apoptotic < Dead. Since it was previously concluded in the section covering the effects of the variation of nutrients on the cell surface glycoprofile (see section 4.9.1), the quantity of lectin binding sites increases as cells go through the DNA cell cycle. This is due to possibly cell size enlargement which was observed from Figure 4.42. This conclusion can be further supported by the fact that the G2/M and S subpopulation curves of each lectin reveal a repetition of Go/G1 curve pattern. Such pattern repetition was also observed in LECTIN-A data of the cells subjected to different levels of nutrients (Figure 4.35).

The bottom plot of the figure allows the identification of the strongest and weakest LECTIN-A signal from the DNA subpopulations. Signal from WGA was the strongest, while LEC A and PNA seemed to be the weakest. Therefore, it can be concluded that the quantity of N-Acetylglucosamine groups available for binding was at a higher number in relation to Fucose, Sialic acid and Galactose, particularly in relation to Galactose as both PNA and LEC A specifically bind to this sugar (see section 4.7).

The pattern of the DNA curves of an in-house lectin and its commercial counterpart was very similar, demonstrating the lectins were interacting with the same sites on the cell surface. However, AAL-2 and WGA showed a very different pattern in response to temperature variation ,unlike that observed from nutrient level variation studies (see section 4.9.1).

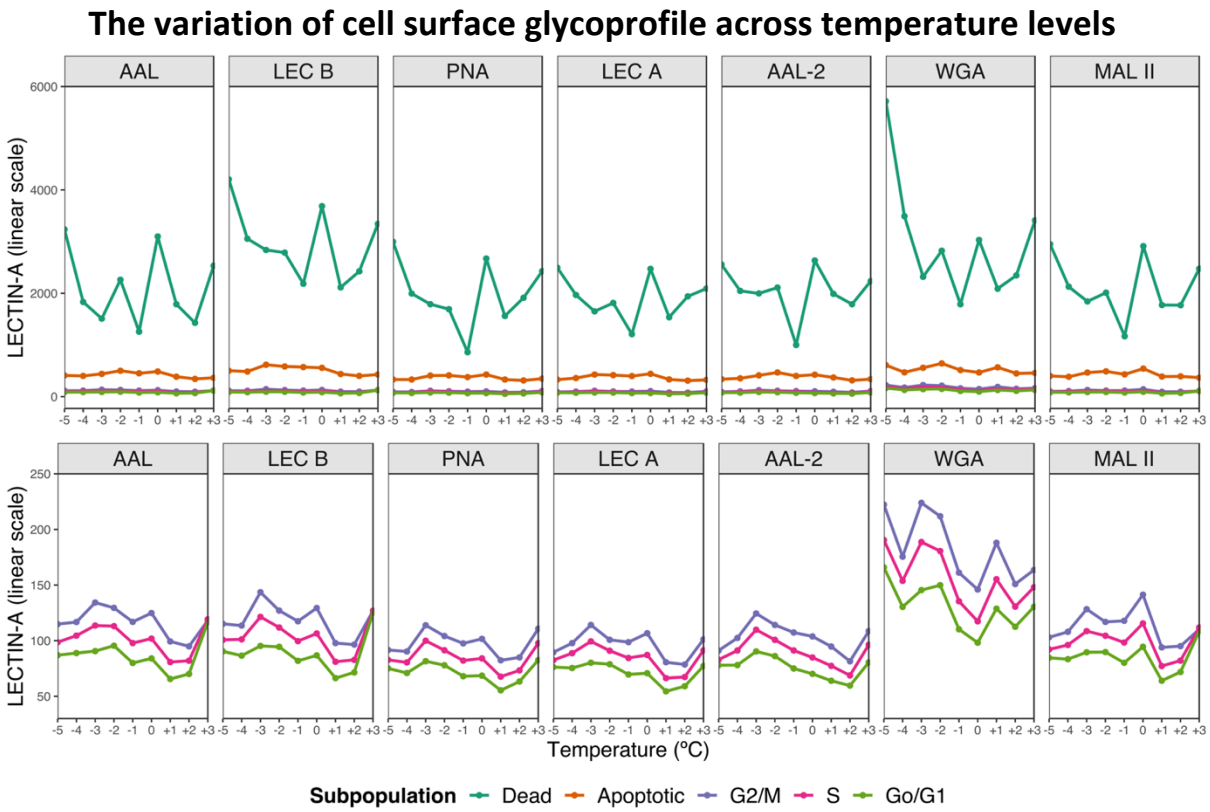

Figure 4.44: Lectin-facetted line plots demonstrating the lectin interaction (LECTIN-A) variation as a function of temperature for 5 different cell subpopulations: dead, apoptotic, and DNA subpopulations (G2/M, S, and Go/G1). The lower plot highlights the DNA subpopulation curves to better visualize their trends. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.8).

Figure 4.45 shows the data distributions of the temperature treatments, including the baseline. However, the baseline boxplot is colored in grey to highlight it as the data distribution to which the different treatments were compared (see section 3.12.7). The investigation of the variation of cell surface glycoprofile of live cells is the main goal of this research work; thus, the DNA subpopulations are the cells of most interest. Therefore, the figure shows data of G2/M, S, and Go/G1 only.

Inferential analysis of the variation of cell surface glycoprofile across temperature levels

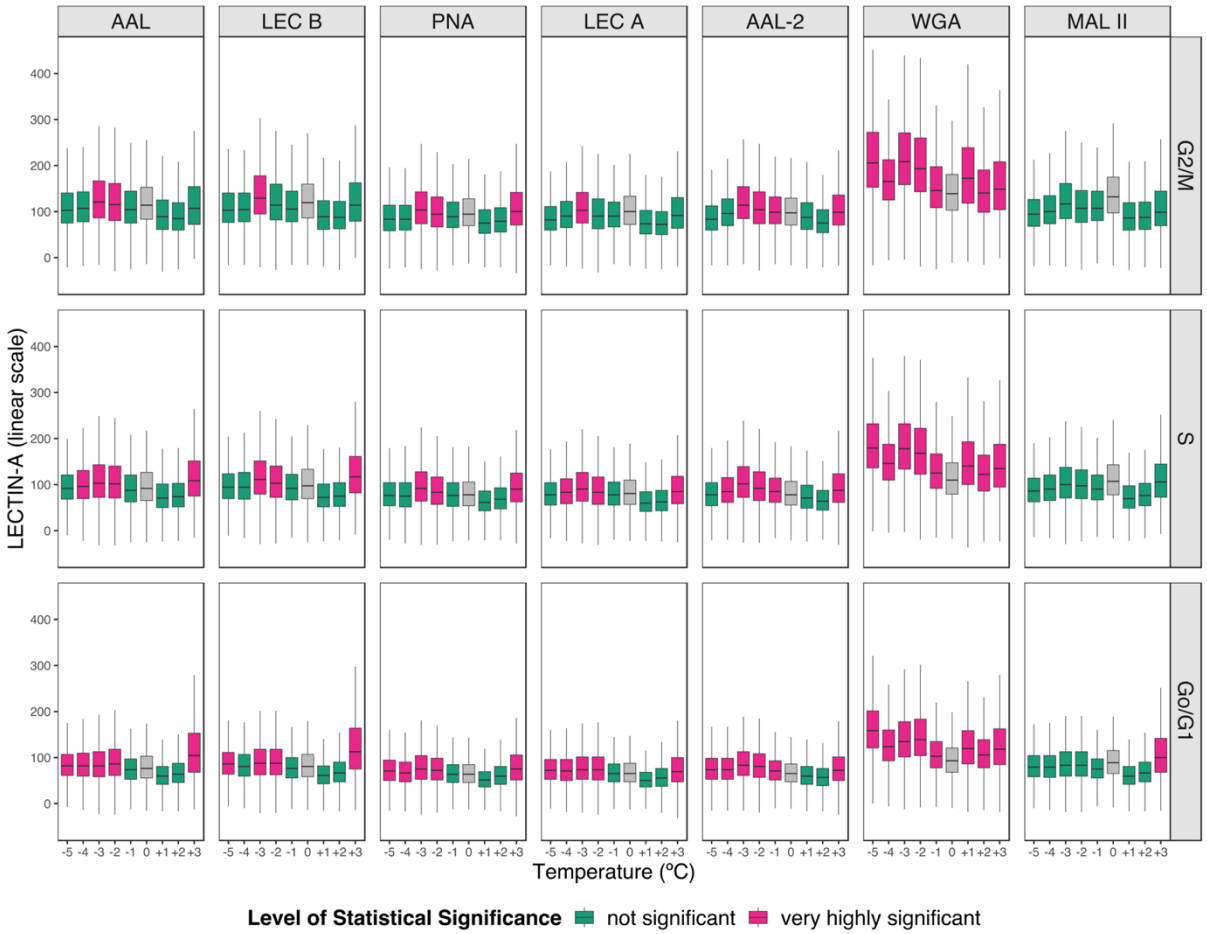

Figure 4.45: Box plot facetted by lectin and DNA subpopulations highlighting the levels of statistical significance of the glycoprofile difference between the temperature treatments and the baseline. Data obtained through a BD FACSAria™ I flow cytometer using LECTIN-A detector channel (see sections 3.12.1.8 and 3.12.7).

Figure 4.45 allows the observation of the highest number of *very highly significant* glycoprofile changes in the Go/G1 subpopulation. As discussed in Section 4.9.1, the sample size of the Go/G1 subpopulation was substantially higher than S and G2/M, thereby increasing the statistical ability to detect significant changes in Go/G1 subpopulations. Table 4.5 summarizes the number of *very highly significant* changes which were detected by each lectin and the temperature treatments in which these changes were found in the Go/G1 subpopulation.

Table 4.5: Table summarizing the number of *very highly significant* changes detected by each lectin and the temperature levels in which these changes were found in the Go/G1 subpopulation.

Lectin	Number of *very highly significant* changes	Treatments in which significant changes were found
WGA	8	-5, -4, -3, -2, -1, +1, +2 and +3°C
AAL-2	6	-5, -4, -3, -2, -1, and +3°C
AAL	5	-5, -4, -3, -2, and +3°C
LEC A	5	-5, -4, -3, -2, and +3°C
PNA	5	-5, -4, -3, -2, and +3°C
LEC B	4	-5, -3, -2, and +3°C
MAL II	1	+3°C

WGA detected *very highly significant* changes in all temperature treatments, while MAL II detected only one change at +3°C. However, most of the lectins was able to detect more than 4 significant changes.

With the purpose of assessing the ability of the methodology to detect scientifically meaningful differences, power analysis was performed. Figure 4.46 shows the results of the analysis covering the DNA subpopulations. It can be observed that the highest power values are in the data obtained from Go/G1 cells. This observation is in agreement with what was observed in Figure 4.45 which showed the highest number of *very highly significant* changes in Go/G1 cells.

Figure 4.47 demonstrates the number of cells used in each treatment to perform the statistical analysis. As previously pointed out, Go/G1 cells were in the highest number, thereby increasing the power computed from this subpopulation. As a consequence, the following discussion is focused on Go/G1 cells.

Power analysis of the glycoprofile changes detected across temperature levels

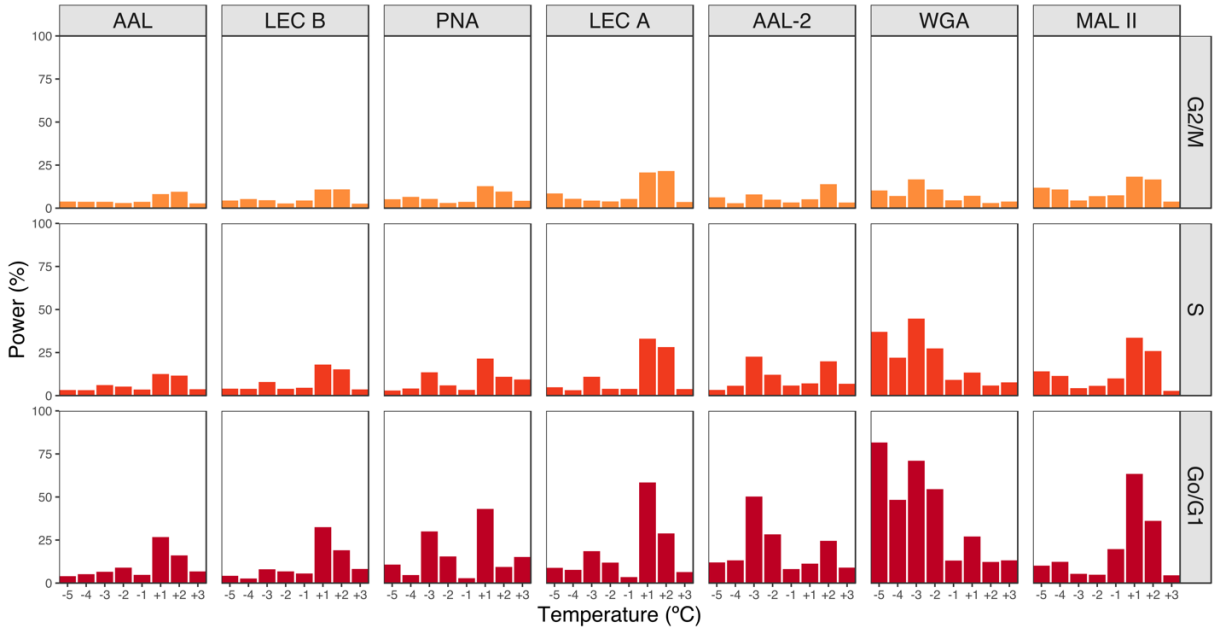

Figure 4.46: Bar plot facetted by lectin and DNA subpopulation demonstrating the results of power analysis of cell surface glycoprofile differences which were detected between the multiple levels of temperature (treatments) and the baseline. Powers analysis on the data obtained through a BD FACSAria™ I flow cytometer (see sections 3.12.1.8 and 3.12.7).

Figure 4.47 also shows a considerable decline in the sample size of treatments in which cells were subjected to increasing temperature levels. This reduction in sample size is a reflection of the sharp drop in cell viability shown in Figure 4.40.

Although sample sizes varied across the different levels of temperature, each Go/G1 sample contained more than 30,000 cells, except for +3°C treatment for all lectins and +2°C treatment for MAL II. Therefore, by observing Figures 4.46 and 4.47, it can be concluded that sample sizes were large enough to identify scientifically meaningful glycoprofile differences. For instance, the power of +1°C treatment for MAL II was 63%.

Consequently, the power results were mostly influenced by the difference between a treatment and the baseline data variabilities as well as by the difference in the means of the

treatment and baseline. With the purpose of discussing the relationship of power, data variability difference and the difference in the means, a more detailed bar plot containing treatment power values and a box plot overlaid with the curve constructed with the means of LECTIN-A signals of each temperature treatment is found in Figure 4.48.

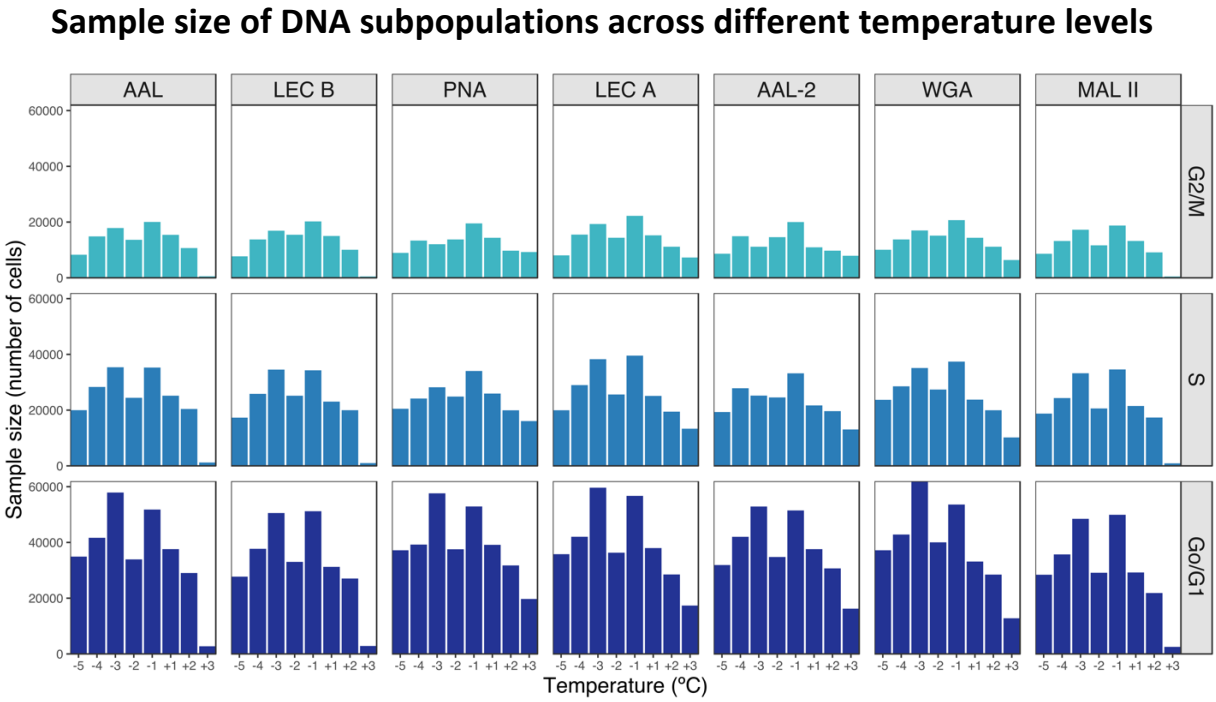

Figure 4.47: Bar plot facetted by lectin and DNA subpopulation demonstrating the sample sizes used in the statistical analysis of cell cultures treated with different levels of temperature. Data obtained through a BD FACSAria™ I flow cytometer (see sections 3.12.1.8 and 3.12.7).

According to Figure 4.48, more scientifically meaningful glycoprofile changes were detected at temperature levels above the baseline level. However, more scientifically meaningful changes were detected at treatment levels below the baseline level by AAL-2 and WGA. In other words, the number of available Fucose, Galactose and Sialic acid sites was altered due to increased temperature levels, whereas the number of N-Acetylglucosamine was altered mostly due to reduced temperature levels.

Although the inferential analysis has classified most of the glycoprofile differences of treatments with low power values as *very highly significant*, descriptive analysis has demonstrated fairly straight curves indicating the absence of a considerable change in the central tendency of the data distributions of these treatments. In addition, the distributions revealed higher data variability in relation to the baseline distribution. For instance, scientifically meaningful change in the number of available Fucose sites on cell surface due to decreased temperature levels is 8% or 9% to the most according to LEC B and AAL, respectively. Therefore, even though *very highly significant* changes were found in the experiments, the chances of these changes to be found are only 9% at the highest as shown by AAL.

It can be concluded that as the temperature level dropped, the number of available N-Acetylglucosamine increases considerably with 82% of power at -5°C level according to WGA. However, AAL-2 detected the most scientifically meaningful change in N-Acetylglucosamine at -3°C level with 51%. Therefore, it can be concluded that WGA was more powerful at detecting those changes. On the other hand, both lectins have shown that the changes in N-Acetylglucosamine due to increased temperature levels were not as scientifically meaningful as the changes detected at the decreased temperature levels.

Changes in the number of available Fucose sites were more scientifically meaningful due to increased temperature levels. However, the power values were 27% and 33% at the highest according to AAL and LEC B, respectively. Descriptive analysis of AAL and LEC B revealed that the Fucose number dropped at the initial increments in temperature, but showed a tendency in increasing the number of Fucose sites at temperature levels higher than +3°C.

Statistical analysis of Go/G1 cell surface glycoprofile across temperature levels

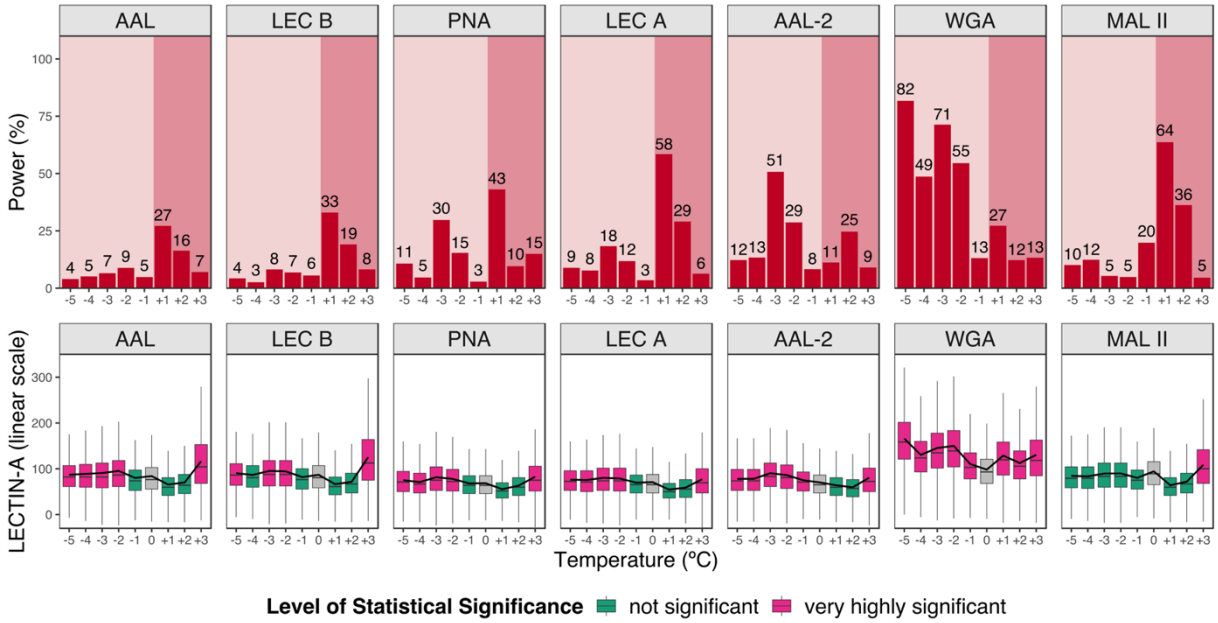

Figure 4.48: Complementary plots summarizing descriptive, inferential and power analysis of temperature treatments applied to Go/G1 cells. **Upper plot:** bar plot facetted by lectin showing power values of each temperature level and highlighting the positive and negative levels using different shades of red. **Lower plot:** Box plot color coded with the results of the inferential analysis and superimposed with a curve constructed by connecting the means of LECTIN-A signal of each temperature level. Data obtained through a BD FACSAria™ I flow cytometer (see sections 3.12.1.8 and 3.12.7).

In the case of the changes in the number of available Sialic acid sites, power values were higher due to increased temperature levels with up to 64%. However, power values of decreased temperature levels were mostly very low showing that the likelihood of finding a scientifically meaningful change due to these treatments was 20% at the highest at -1°C. In fact, descriptive analysis revealed a fairly straight curve in those treatments indicating that the central tendency of the data distributions did not change considerably. In other words, the number of available Sialic acid sites did not change in response to reduced levels of temperature, while initial increased temperature levels caused the reduction in the number of Sialic sites and a tendency to increase this number to the same level as the baseline was shown at +3°C

temperature treatment. It is known that glycosylation changes in increased temperature (heat shock) can be associated with the reduction in sialylation due to increasing sialidase activity in the supernatant resulted from the release of proteases of the increasing number of dead cells (Clark, Chaplin and Harcum, 2004).

Many studies have reported alterations in the host cell proteome owing to sub-physiological temperature in the cell culture process (Baik *et al.*, 2006; Underhill and Smales, 2007; Dietmair *et al.*, 2012). Consequently, there is a growing number of scientific studies demonstrating the relationship between mild temperature decrease and the glycosylation profile of therapeutic proteins (Clark, Chaplin and Harcum, 2004; Bollati-Fogolín *et al.*, 2005; Woo *et al.*, 2008; Sou *et al.*, 2015).

Since the cellular glycosylation machinery is the same for all synthesized proteins, the findings of those studies on glycosylation alterations are likely to be correlated with the glycosylation changes observed on the cell surface. A study investigated glycosylation changes in a monoclonal antibody (mAb) expressed in CHO-T cells cultured at 36.5°C and with a temperature shift to 32°C during late exponential/early stationary phase and a decrease in the proportion of the more processed glycan structures on the constant region of the mAb was demonstrated. The levels of mRNA expression of these glycosyltransferase enzymes were measured: one N-acetylglucosaminyltransferase (GnTII), two galactosyltransferases (β-GalTI and β-GalTIII), and a fucosyltransferase (FucT). It was shown that the mRNA expression levels of these enzymes were considerably lower at 32°C (Sou *et al.*, 2015). However, this present research work has found no scientifically meaningful changes in Fucose and Galactose, but a meaningful increase in N-Acetylglucosamine in 32°C of CHO-K1 cell cultures. Nonetheless, the cellular responses to cold shock vary between cell lines, expression systems and product of interest (Al-Fageeh and Smales, 2006). For example, another study with an rCHO cell line has found that profiles of antennary structures and N-linked glycan of Erythropoietin expressed at 32°C and 38°C were comparable (Woo *et al.*, 2008).

In summary, as increased levels of temperature compromised cell viability and raised the medium pH, it can be concluded that an increase in the number of available Fucose, Galactose, N-Acetylglucosamine and Sialic acid sites on CHO-K1 cell surface are glycoprofile alterations

associated with temperature levels above 37°C. Thus, such increase can be an indicative of modifications in the cellular metabolism associated with harmful levels of temperature. An increase in the cell internal and external complexity level and in the relative cell size are also changes associated with temperature levels above 37°C.

4.9.3 Analysis of the effects of CO_2 variation

The experimental data involving the variation of CO_2 comprises 10 datasets each representing a CO_2 level: 9 datasets consist of the variation of CO_2 level in the last 24 hours of culture, and 1 dataset extracted from cells grown at CO_2 baseline condition throughout the entire 96 hours of culture (sample collection point).

In order to facilitate the understanding of the data, the variables of interest were plotted against the variation of CO_2 in relation to the setpoint or the baseline condition (5% of CO_2). In other words, the 0 point of the x axis (Level of carbon dioxide) depicts the data obtained from cells cultivated at baseline conditions, whereas the remaining points depict the datasets obtained from the different levels of CO_2 measured in one unit of a percentage. The experimental setup is fully described in section 3.12.2.3.

Cell culture parameters

Figure 4.49 allows the observation of cell viability alteration in response to the variation of CO_2 levels. Up to 0 (baseline CO_2 level), cell viability of all lectin samples was within 85 and 92%. However, a sharp increase in cell viability took place between 0 and +1% reaching the viability range of 93 and 97%. Between +1 and +3%, cell viability was stable, but a dramatic decrease occurred due to further increases in CO_2 levels and the lowest percentage of viable cells at +5% was 80% for MAL II samples.

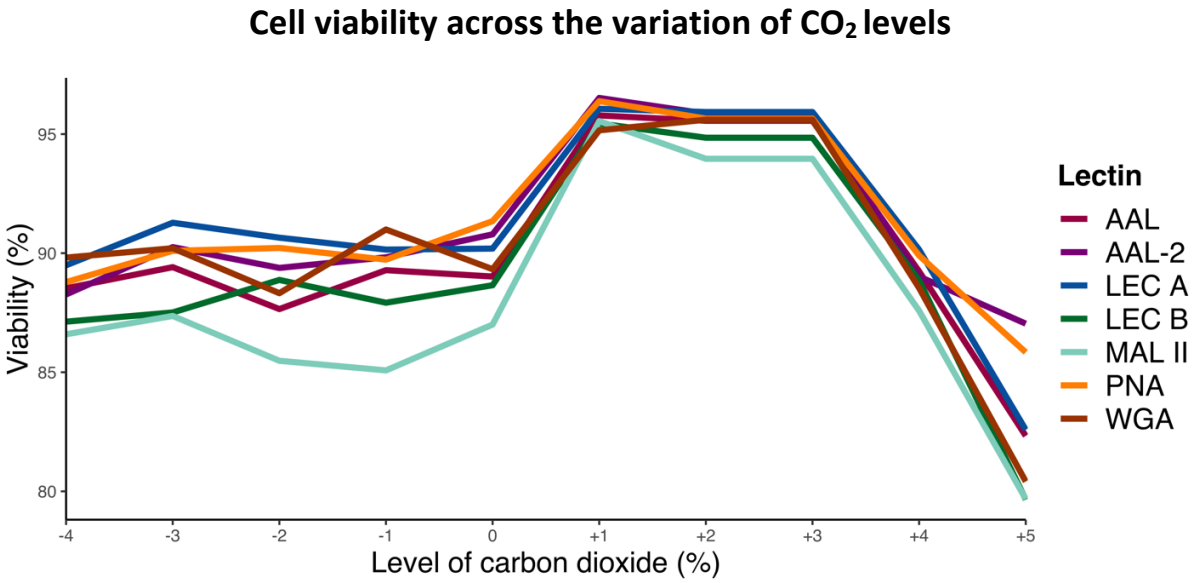

Figure 4.49: Line plot showing polynomial fits of cell viability of all lectin samples from the experiment involving the variation of CO_2 levels. Data obtained through a BD FACSAria™ I flow cytometer and using 7-AAD to stain dead cells, thus determining cell viability (see section 3.12.1.6).

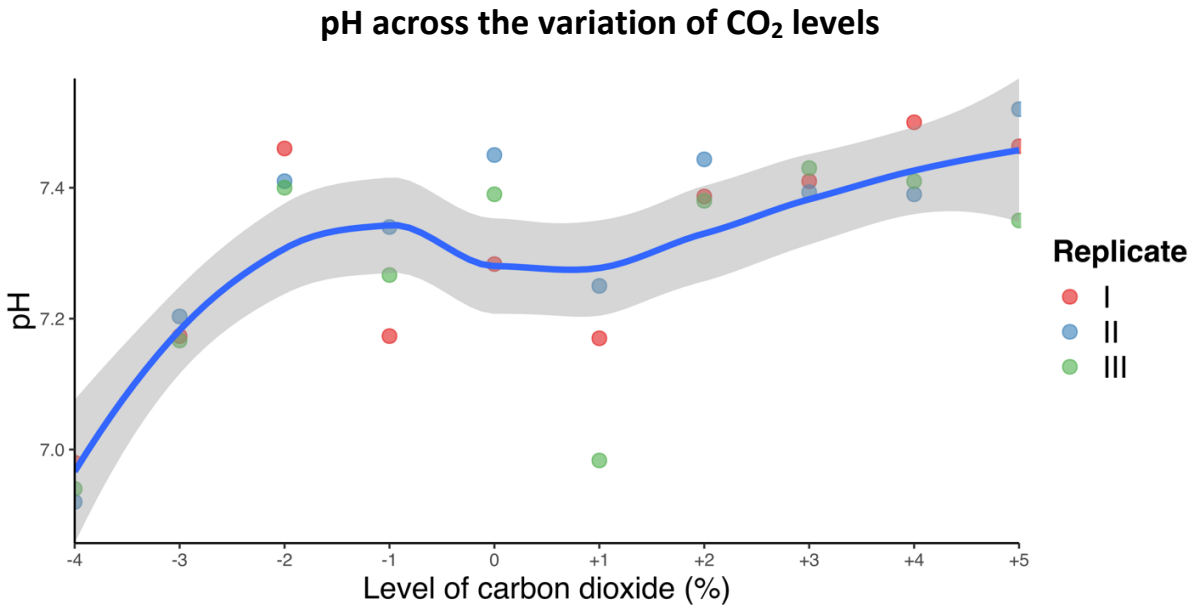

Figure 4.50: Line plot showing a polynomial fit with 95% confidence interval of pH as a function of the variation of CO_2 levels. The values of pH were obtained using an Orion Semi-micro pH electrode (see section 3.12.1.8).

Figure 4.50 shows the change in pH as the level of CO_2 was varied. It can be observed that CO_2 levels below the baseline level reduced the pH of the cell cultures lower than 7 at -4%. On the other hand, levels above the baseline caused an increase in the pH from 7.3 to higher than 7.5 at +5%. Therefore, the increase was not considerable and the first increment in CO_2 level (from 0 to +1%) did not cause an increase in the pH. Additionally, as previously observed in Figures 4.32 and 4.41, the rise in pH is associated with the decrease in cell viability.

Descriptive analysis of the variation of the relative cell size and cell internal and external complexity parameters

Figure 4.51 demonstrates the variation of the relative cell size means as cells were subjected to a different level of CO_2 in the last 24 hours of the cell culture process. For all lectins, it can be clearly seen that the pattern of the change in FSC-A means was the same for all subpopulation curves. The relative cell size of the subpopulations increased in response to both first increase and decrease in the baseline CO_2 level (5%). However, a sharp decrease in FSC-A signal was detected between -3 and -4%. Generally, for CO_2 levels above the baseline, there was an increase in the relative cell size which stopped between +1 and +3% followed by a decrease.

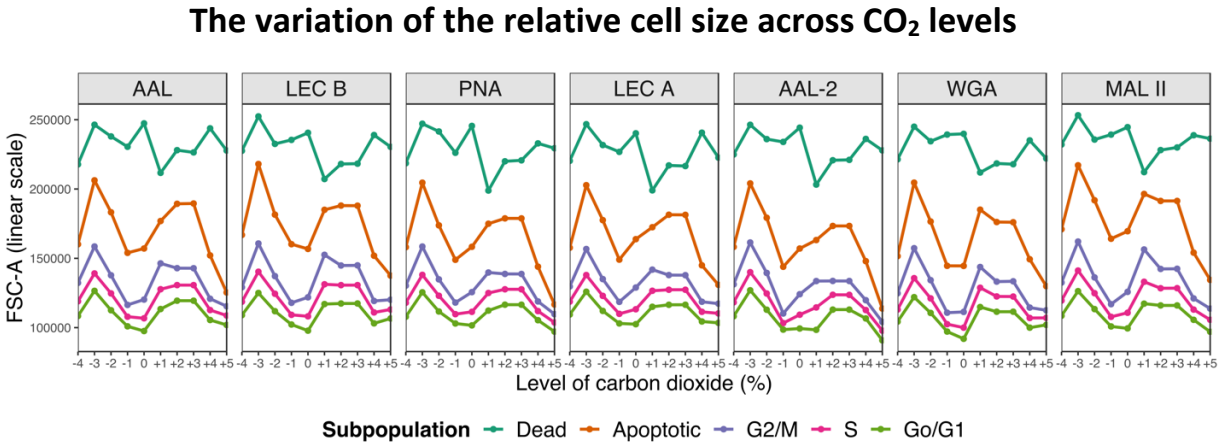

Figure 4.51: Lectin-facetted line plot showing the relative cell size (FSC-A) variation as a function of CO_2 for 5 different cell subpopulations: dead, apoptotic, and DNA subpopulations (G2/M, S, and Go/G1). Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.8).

Likewise as previously observed from FSC-A datasets of nutrient and temperature variation experiments (Figures 4.33 and 4.42), the positions of the relative cell size curves in Figure 4.51 in relation to FSC-A scale were slightly different across the lectins, demonstrating that the lectin interaction on the surface of the cell might not significantly influence the FSC-A signal or all lectins from the panel influence the signal to the same degree. Another similarity observed between the aforementioned FSC-A datasets and CO_2 FSC-A dataset, is the cell size order across the subpopulations: Go/G1 < S < G2/M < Apoptotic < Dead.

The variation of the relative cell internal and external complexity across different CO_2 levels

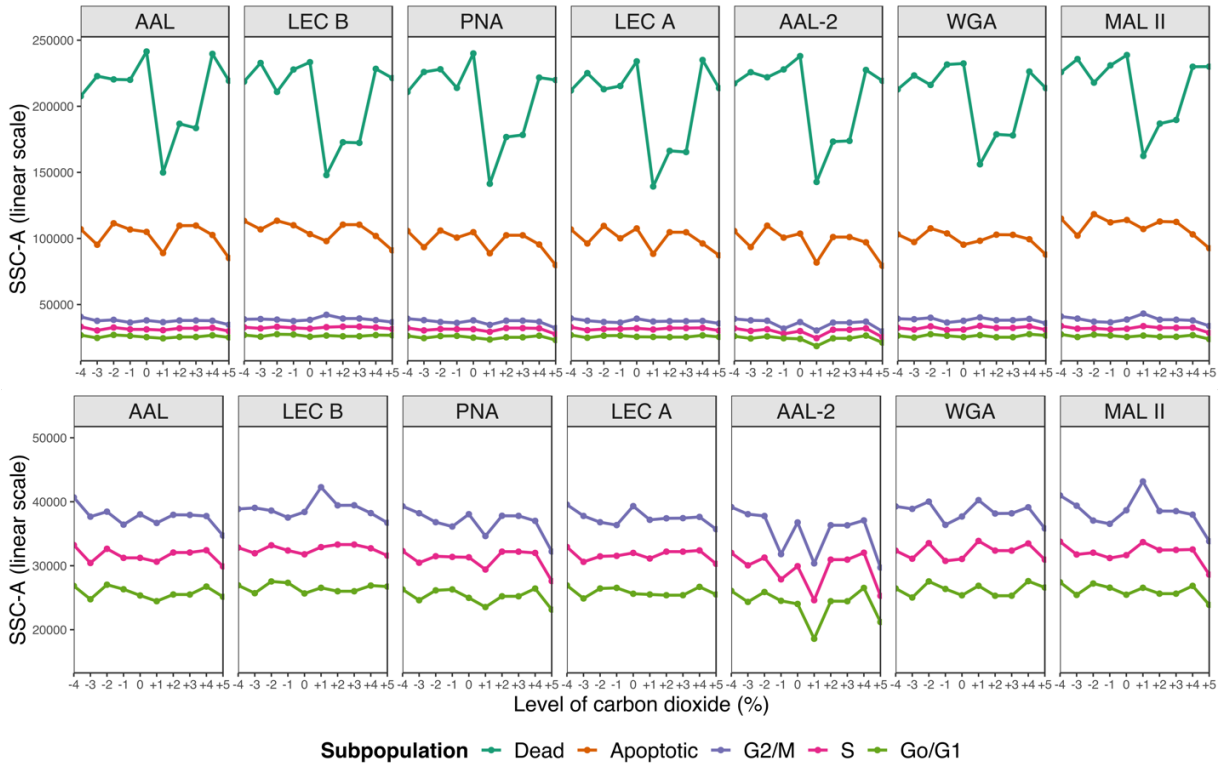

Figure 4.52: Lectin-facetted line plots demonstrating the relative cell internal and external complexity (SSC-A) variation as a function of CO_2 level for 5 different cell subpopulations: dead, apoptotic, and DNA subpopulations (G2/M, S, and Go/G1). The lower plot highlights the DNA subpopulation curves to better visualize their trends. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.8).

Figure 4.52 shows the alteration of the relative cell internal and external complexity parameters as the CO_2 level was varied. The upper plot shows 5 subpopulation curves for each lectin, while the lower plot highlights the curves of the DNA subpopulations, facilitating the observation of the trends revealed by these curves. Overall, all lectin curves showed a very similar pattern, particularly the DNA curves which are the subpopulations of most interest since they were composed of live cells. However, AAL-2 DNA curves demonstrated a different pattern. A considerable change in SSC-A signal was not observed in either below or above the baseline CO_2 level (0 point in the x axis).

Likewise Figures 4.34 and 4.43, the order Go/G1 < S < G2/M < Apoptotic < Dead was observed in relation to SSC-A scale in Figure 4.52, validating once more the previous conclusion on the increasing cellular metabolism level as cells go through the DNA cell cycle.

Statistical analysis of cell surface glycoprofile variation

The workflow of this section is the same adopted previously for the statistical analysis of the variation of cell surface glycoprofile in response to different levels of nutrients and temperature. Therefore, the investigative analysis of the cell surface glycoprofile variation due to CO_2 alteration initiates with a descriptive analysis of data obtained from LECTIN-A detector channel, then the inferential and power analysis of the data is developed.

Figure 4.53 demonstrates the variation of LECTIN-A signal as the level of CO_2 changed, allowing the investigation of cell surface glycoprofile variation. As previously observed on the data from FSC-A and SSC-A detector channels, the order of the LECTIN-A signal intensity of the subpopulations of all lectins is: Go/G1 < S < G2/M < Apoptotic < Dead. The figure allows the observation of the increasing LECTIN-A signal (lectin binding) as cells go through the DNA cell cycle; such conclusion was also drawn from the datasets of LECTIN-A of nutrient and temperature variation (Figures 4.35 and 4.44). This is due to possibly cell size enlargement as observed in the Figure 4.51. This conclusion can be further supported by the fact that the G2/M and S subpopulation curves of each lectin revealed a repetition of Go/G1 curve pattern. Such pattern repetition was also observed in LECTIN-A datasets of cells subjected to different levels of nutrients and temperature (Figures 4.35 and 4.44).

As can be observed from the bottom plot of Figure 4.53, WGA was the strongest LECTIN-A signal, while PNA, LEC A and AAL-2 were the weakest signal. Thus, it can be concluded that the quantity of N-Acetylglucosamine groups available for binding was at a higher number in relation to Fucose, Sialic acid and Galactose, particularly in relation to Galactose as both PNA and LEC A specifically bind to this glycan (see section 4.7).

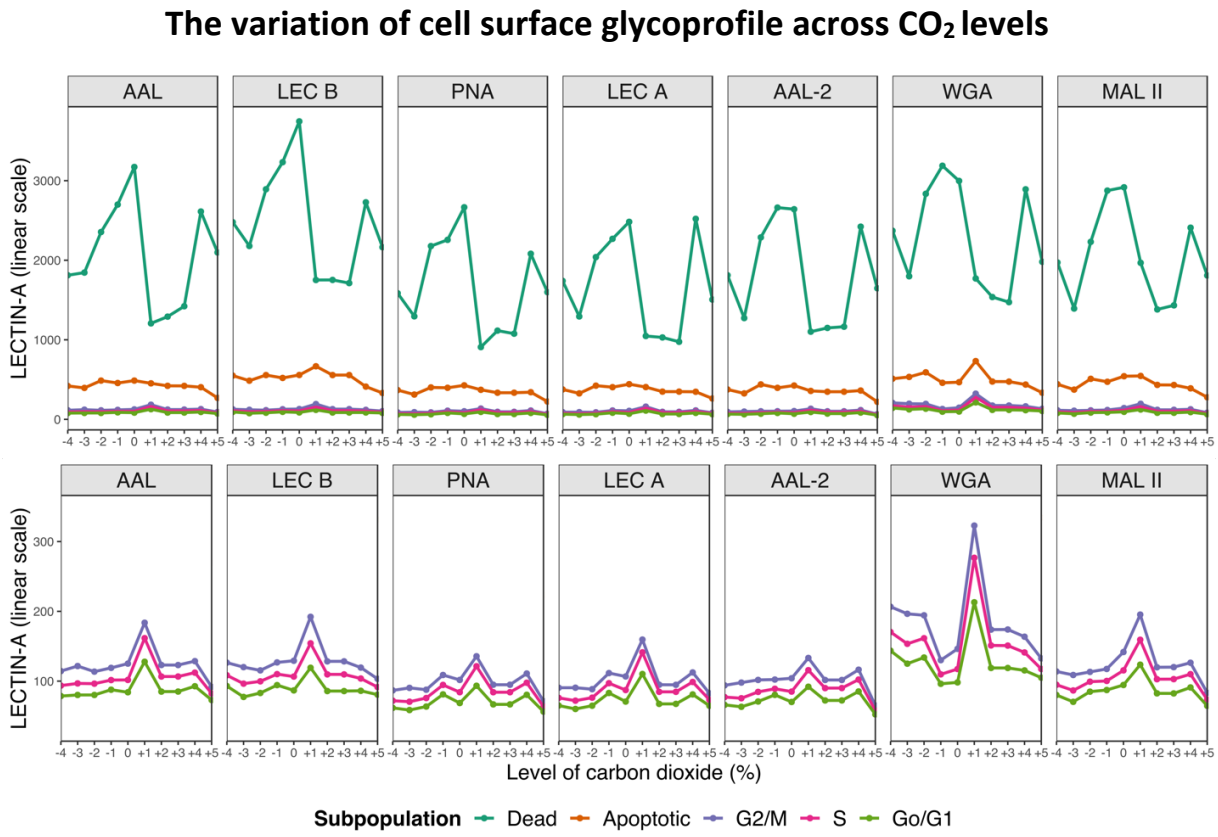

Figure 4.53: Lectin-facetted line plots demonstrating the lectin interaction (LECTIN-A) variation as a function of CO_2 for 5 different cell subpopulations: dead, apoptotic, and DNA subpopulations (G2/M, S, and Go/G1). The lower plot highlights the DNA subpopulation curves to better visualize their trends. Data obtained through a BD FACSAria™ I flow cytometer (see section 3.12.1.8).

The patterns of the DNA curves of an in-house lectin and its commercial counterpart were very similar, demonstrating the lectins were interacting with the same sites on cell surface. However, AAL-2 and WGA have shown different patterns in response to CO_2 variation in contrast to the data obtained from nutrient level variation studies (see section 4.9.1). Nevertheless, the pattern difference between AAL-2 and WGA was also observed in LECTIN-A data of temperature variation studies (Figure 4.44).

Figure 4.54 shows the data distributions of the CO_2 treatments, including the baseline. However, the baseline boxplot is colored in grey to highlight it as the data distribution to which

the different treatments were compared (see section 3.12.7). The investigation of the variation of cell surface glycoprofile of live cells is the main goal of this research work; therefore, the DNA subpopulations contained the cells of most interest. Thus, the figure shows only the data of G2/M, S, and Go/G1.

Figure 4.54 allows the observation of the highest number of *highly* and *very highly significant* changes detected in the Go/G1 subpopulation, similar to Figures 4.36 and 4.45. As previously discussed, the sample size of Go/G1 subpopulation was substantially higher than S and G2/M, thereby increasing the statistical ability to detect significant changes in Go/G1 subpopulations. As a result, the following analysis is focused on Go/G1; thus, Table 4.6 summarizes the number of *highly* and *very highly significant* changes which were detected by each lectin and the CO_2 treatments in which these changes were found in Go/G1 subpopulation.

Inferential analysis of the cell surface glycoprofile variation across CO_2 levels

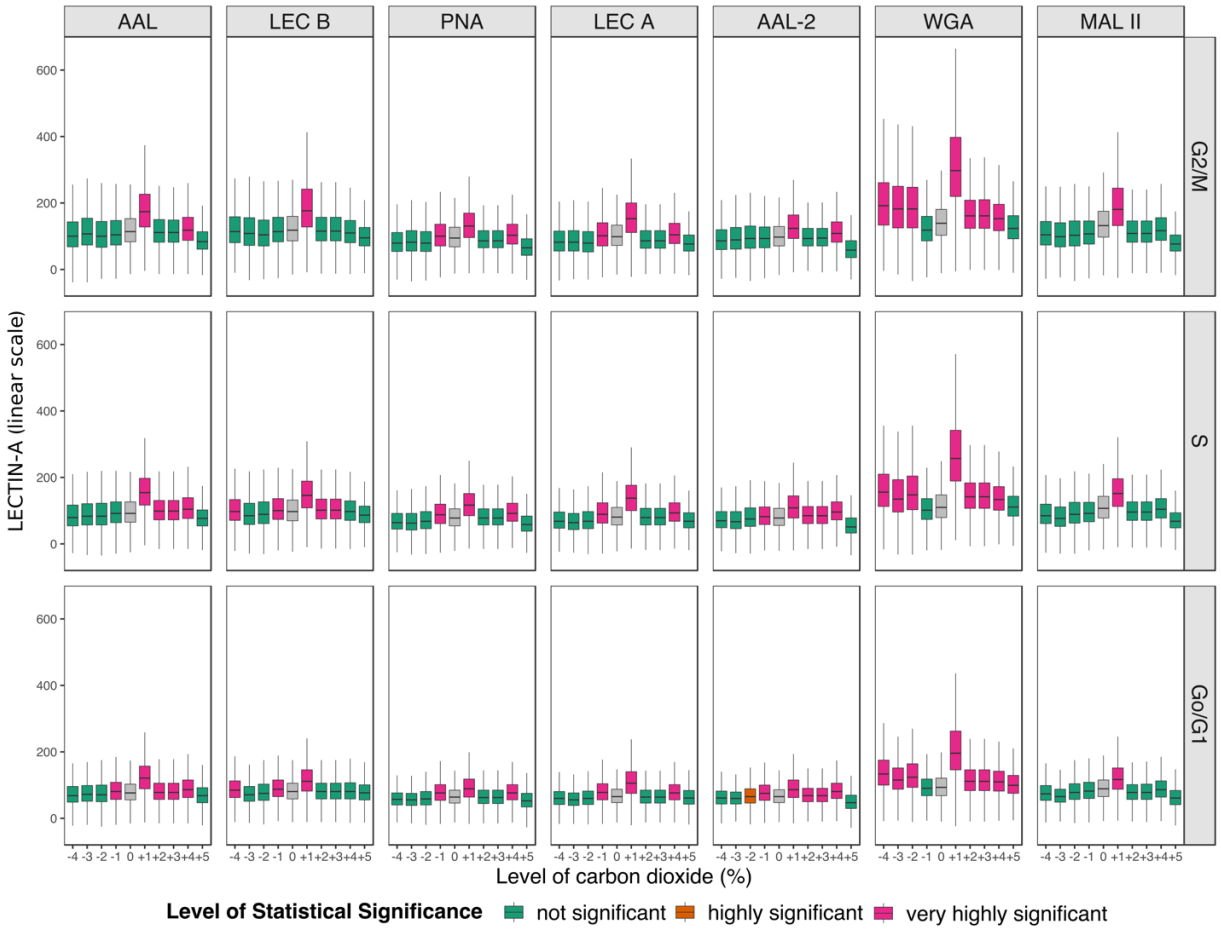

Figure 4.54: Box plot facetted by lectin and DNA subpopulations highlighting the levels of statistical significance of the glycoprofile difference between the CO_2 treatments and the baseline. Data obtained through a BD FACSAria™ I flow cytometer using LECTIN-A detector channel (see sections 3.12.1.8 and 3.12.7).

Table 4.6: Table summarizing the number of *highly and very highly significant* changes detected by each lectin and the CO_2 levels in which these changes were found in the Go/G1 subpopulation.

Lectin	Number of *highly and very highly significant* changes	Treatments in which significant changes were found
WGA	8	-4, -3, -2, +1, +2, +3, +4, and +5%
AAL-2	6	-2, -1, +1, +2, +3, and +4%
AAL	5	-1, +1, +2, +3, and +4%
LEC B	3	-4, -1, and +1%
PNA	3	-1, +1, and +4%
LEC A	3	-1, +1, and +4%
MAL II	1	+1%

WGA detected *very highly significant* changes in 8 out of 9 CO_2 treatments, while MAL II detected only one at +1%. However, most of lectins was able to detect three or more significant changes.

The power of the changes detected was computed to evaluate the likelihood of finding a significant difference when there was one. Figure 4.55 shows the results of the analysis covering the DNA subpopulations. It can be observed that the highest power values are in the data obtained from Go/G1 cells. This observation is in agreement with what was observed in Figure 4.54 that showed the highest number of *highly* and *very highly significant* changes in Go/G1 cells.

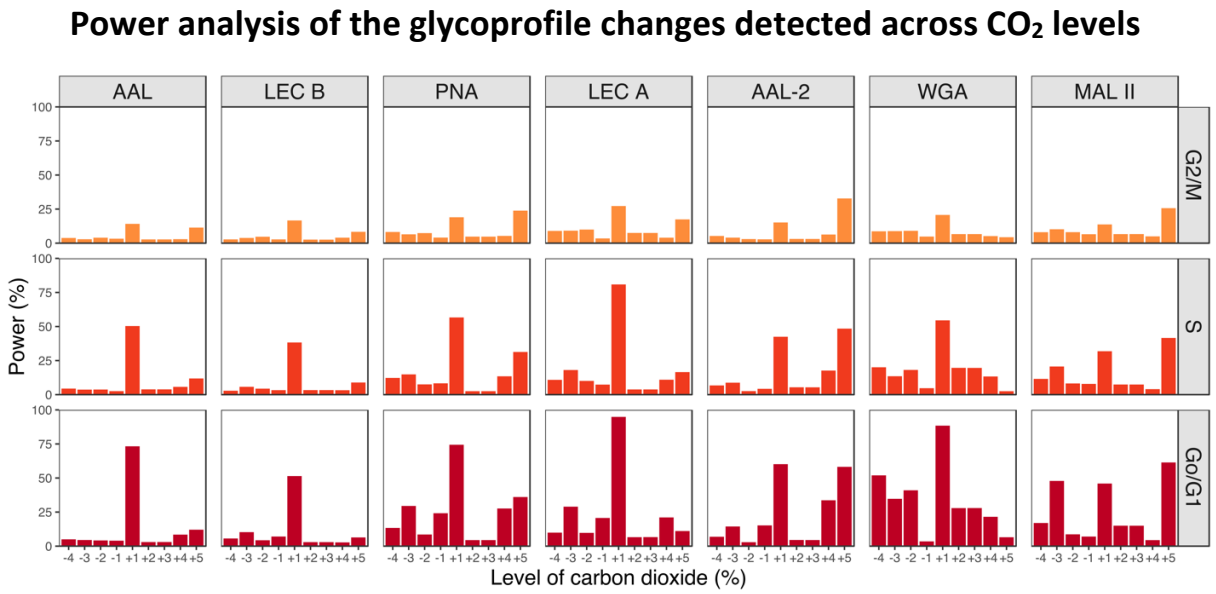

Figure 4.55: Bar plot facetted by lectin and DNA subpopulation demonstrating the results of power analysis of the cell surface glycoprofile differences which were detected between the multiple levels of CO_2 (treatments) and the baseline. Powers analysis on the data obtained through a BD FACSAria™ I flow cytometer (see sections 3.12.1.8 and 3.12.7).

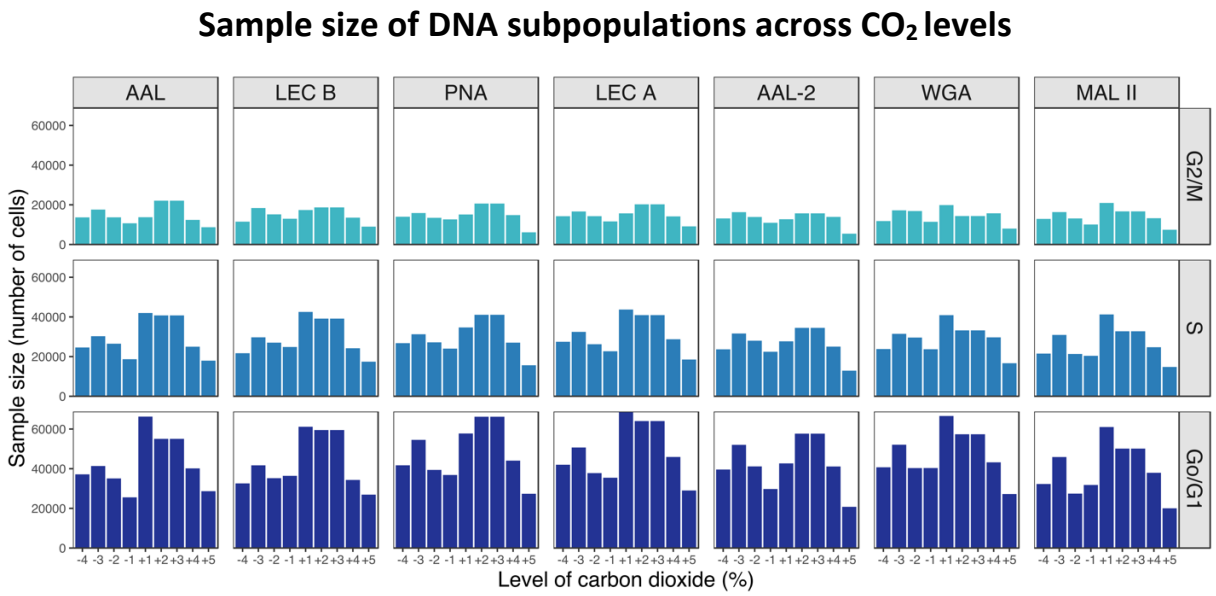

Figure 4.56: Bar plot facetted by lectin and DNA subpopulation demonstrating the sample sizes used in the statistical analysis of cell cultures treated with different levels of CO_2. Data obtained through a BD FACSAria™ I flow cytometer (see sections 3.12.1.8 and 3.12.7).

Figure 4.56 above demonstrates the number of cells used in each treatment to perform the statistical analysis. As previously pointed out, Go/G1 cells were in the highest number thereby increasing the power computed from this subpopulation. As a result, the following discussion is focused on Go/G1 cells. The figure also shows a decline in the sample size of treatments in which cells were subjected to increasing CO_2 levels. This reduction in sample size is a reflection of the sharp drop in cell viability shown in Figure 4.49.

Although sample sizes varied across the levels of CO_2, each Go/G1 sample contained more than or equal to 20,000 cells. Therefore, by observing Figures 4.55 and 4.56, it can be concluded that sample sizes were large enough to identify scientifically meaningful glycoprofile differences. For instance, although sample sizes were just above 20,000 cells, power of +5% treatment for AAL-2 was 58% and 61% for MAL II.

Consequently, the power results were mostly influenced by the difference between a treatment and the baseline data variabilities as well as by the difference in the means of the treatment and baseline. With the purpose of discussing the relationship of power, data variability difference and the difference in the means, a more detailed bar plot containing treatment power values and a box plot overlaid with the curve constructed with the means of LECTIN-A signals of each CO_2 treatment is found in Figure 4.57.

Generally, according to Figure 4.57, more scientifically meaningful glycoprofile changes were detected at CO_2 levels above the baseline level. By observing the data in the figure, it can be concluded that the number of Fucose available sites on the cell surface did not change significantly due to the reduction or increase in CO_2 baseline level. However, +1% treatment caused a scientifically meaningful change in the number of Fucose sites with 73% power computed from AAL data and 51% from LEC B. In fact, this CO_2 level treatment resulted in a scientifically meaningful increase in LECTIN-A signal from all lectins. PNA power values was 74%, LEC A was 95%, AAL-2, WGA and MAL II, was 60%, 88% and 46%, respectively. Therefore, the increase of +1% of CO_2 in the baseline level was demonstrated to cause a considerable increase in the number of Fucose, Galactose, N-Acetylglucosamine and Sialic acid sites on cell surface. However, a further increase in CO_2 level caused a decrease in the number of the aforementioned sugar molecules to a point which the glycoprofile difference in relation to the

baseline is very unlikely to be detected. At +2% CO_2 level, WGA power dropped from 88% to 28%, MAL II from 46% to 15%, AAL-2 from 60% to 5%, AAL from 73% to 3%, LEC B from 51% to 3%, and the most dramatic reduction in power was computed from LEC A data which showed a reduction from 95% to 7% at +2% CO_2 level. The high power levels computed at +1% CO_2 treatment, could, perhaps, be partly due to the fact that the sample sizes were larger for this treatment (Figure 4.56). However, the bottom plot in Figure 4.57 clearly shows a considerable increase in LECTIN-A signal for all lectins.

The power profiles of PNA and LEC A were very similar, showing that the number of Galactose sites on cell surface were very likely to rise due to one level of increase in the CO_2 baseline level (at +1% CO_2 treatment), but the remaining CO_2 levels showed low scientifically meaningful impact on Galactose sites.

Scientifically meaningful changes in the number of N-Acetylglucosamine sites were more likely to be found from WGA signal. It can be concluded that the number of N-Acetylglucosamine is likely to rise due to decreasing CO_2 levels. However, following a sharp increase in N-Acetylglucosamine caused by +1% CO_2 level (88% power), further increases in CO_2 reduced the number of N-Acetylglucosamine sites to the baseline level number and the likelihood of finding scientifically meaningful glycoprofile changes dropped significantly from 28% to 6% (+2 and +5 CO_2).

In the case of the changes in the number of available Sialic acid sites, power values were higher due to increased CO_2 levels with up to 61%. On the other hand, power values of decreased CO_2 levels were lower showing that the likelihood of finding a scientifically meaningful change due to these treatments was 48% at the highest at -3% level. Descriptive analysis revealed that the number of Sialic acid sites decreased in response to CO_2 level alteration except for +1% CO_2 level. However, the decrease in the number of Sialic acid sites on cell surface was more scientifically meaningful when CO_2 level was increased.

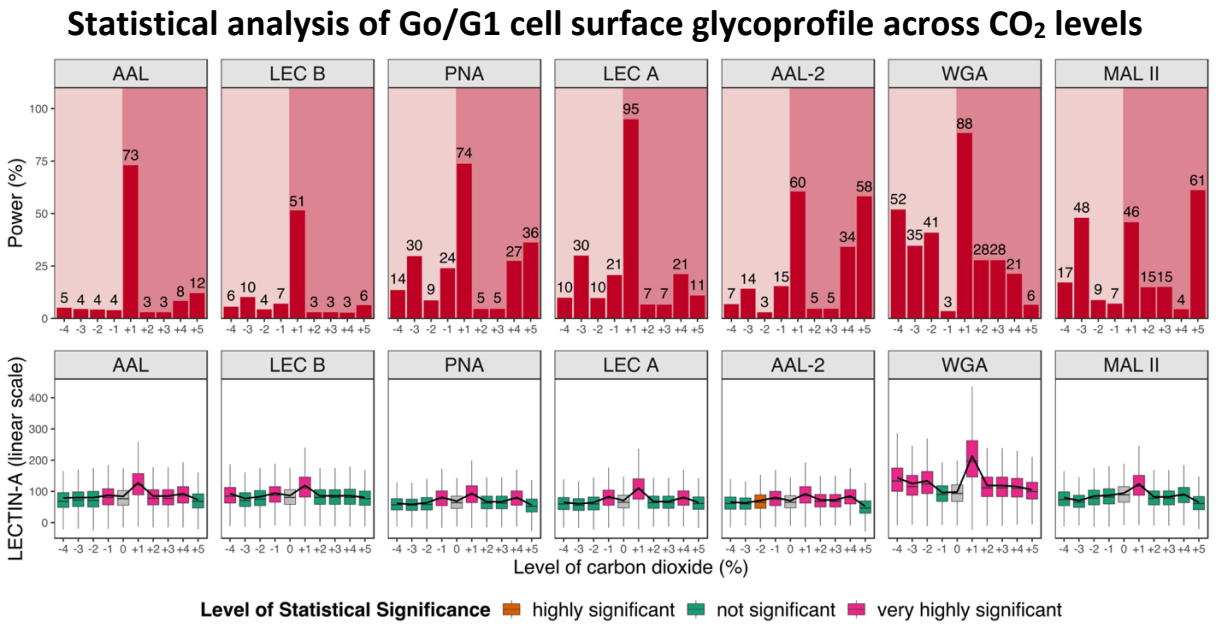

Figure 4.57: Complementary plots summarizing descriptive, inferential and power analysis of CO_2 treatments applied to Go/G1 cells. **Upper plot:** bar plot facetted by lectin showing power values of each CO_2 level and highlighting the positive and negative levels using different shades of red. **Lower plot:** Box plot color coded with the results of the inferential analysis and superposed with a curve constructed by connecting the means of LECTIN-A signal of each CO_2 level. Data obtained through a BD FACSAria™ I flow cytometer (see sections 3.12.1.8 and 3.12.7).

Power results demonstrated that the cell surface glycoprofile responses to increased levels of CO_2 are more scientifically meaningful than the responses to decreased levels. High pCO_2 levels are potentialy associated with glycosylation changes due to the increase in pH resulted from the acidification of the medium. The cellular internal pH can therefore be affected, and particularly the pH of the Endoplasmic reticulum and the Golgi apparatus organelles (Thorens and Vassalli, 1986; Boron, 1987; McQueen and Bailey, 1990). A reduction of polysialic acid content on the cell surface of CHO MT2-1-8 cells due to high levels of pCO_2 has been demonstrated through a flow cytometric analysis. The 5a5 MAb (mouse IgM) primary antibody was used to measure cell surface polysialic acid content which was shown to reduce with increasing pCO_2 in a dose-dependent manner (Zanghi *et al.*, 1999). Another study with the CHO MT2-1-8 cell line demonstrated glycosylation changes in the expressed protein tissue

plasminogen activator (tPA) in response to increased pCO_2 at constant or elevated osmolality. A decrease in the proportion of sialic acids consisting of N-glycolylneuraminic acid was shown in the tPA proteins expressed at 250 mmHg pCO_2 in comparison with the proteins produced at 36 mmHg pCO_2. Additionally, the study demonstrated a decrease in Fucose, N-Acetylglucosamine, Galactose, and Mannose due to high levels of pCO_2 (Kimura and Miller, 1997).

To conclude, as cell viability was compromised and pH of the cell culture rose with increased levels of CO_2, it can be concluded that a decrease in the number of available N-Acetylglucosamine and Sialic acid sites on CHO-K1 cell surface can be indicative of modifications in the cell metabolism associated with CO_2 levels above 5%. In addition, a decrease in the cell internal and external complexity level and a decrease in the relative cell size are also changes associated with CO_2 levels higher than 5%.

4.9.4 Cell surface glycosylation variation: summary and the early detection of the changes in Go/G1 cell population

The three previous sections described in detail the changes on cell surface glycosylation as the cell culture parameters were varied. This section aims to summarize the results to highlight relationships between the overall changes in cell surface glycosylation which were observed by altering each parameter.

At depleted levels of nutrients (negative spent medium levels), GlcNAcylated glycoforms increased while the galactosylated as well as sialylated glycoforms decreased. Such pattern of binding was also observed in positive levels of temperature (temperature levels above 37° C) and both depleted nutrient and increased temperature levels led to the decrease in cell viability (see Figure 4.39 in section 4.9.1, and Figure 4.48 in section 4.9.2). However, the pattern was not observed with excess nutrient and decreased temperature levels in which cell viability was not affected. In fact, viability increased in the case of excess of nutrients (positive spent medium levels). Therefore, the mechanism, which regulates the changes in cell surface

glycosylation when cells are subjected to stressful conditions (which severely affects viability) may be different to the mechanism that alters the cell surface glycosylation without affecting the viability. However, although increased levels of CO_2 also caused a decrease in cell viability, the lectin binding pattern previously described was not as scientifically meaningful (lower power values) as in the depleted nutrient and increased temperature levels (Figure 4.57 in section 4.9.3).

The earliest and most scientifically meaningful cell surface glycosylation change in cells subjected to depleted nutrient levels was a drop in fucosylated glycoforms (83% at -1 day by LEC B), whereas an increase in GlcNAcylated glycoforms was observed in cells subjected to excess nutrient levels (50% at +1 day by AAL-2) (see Figure 4.39 in section 4.9.1). In the case of cells subjected to increased temperature levels, variation in sialylated glycoforms was the earliest meaningful change with a decrease at 38° C (64%). Whereas within decreased temperature levels, a meaningful increase in GlcNAcylated glycoforms was observed at 35° C (55%) (Figure 4.48 in section 4.9.2). Finally, at decreased CO_2 levels, the earliest change was a drop in sialylated glycoforms at 2% of CO_2 (48%), while a rise in galactosylated glycoforms at 6% of CO_2 (95%) was the earliest meaningful change in increased CO_2 levels (Figure 4.57 in section 4.9.3).

4.9.5 Comparative power analysis of the responses of Go/G1 cell surface glycoprofile to process parameter alterations

While the first sections of this chapter covered in detail the variation of cell surface glycoprofile in response to changes in the cell culture parameters, this section aims to establish a general comparative analysis of the effects of these parameters on CHO-K1 cell surface glycoprofile. Since power analysis provides the likelihood of finding scientifically meaningful changes, the results of this analysis can be used to have a comparative insight indicating the cell culture parameter which causes the most meaningful cell surface glycoprofile alterations. In addition, the analysis allows the identification of key lectins and glycans associated with these meaningful alterations.

Therefore, power values of the Go/G1 subpopulation were averaged by each cell culture parameter to obtain a comparative power analysis among these parameters. For instance, power values computed from each lectin dataset of a temperature level (treatment) were summed up and divided by the number of power values, which is 56 in this case. Similarly, power values of each lectin dataset of a cell culture parameter were averaged across the number of levels of the parameter (8 levels for temperature variation) to obtain a power comparative analysis among the lectins and glycans involved.

It can be observed from Figure 4.58 that the variation of the level of spent medium can cause more meaningful glycoprofile alterations on cell surface than the variation of temperature and CO_2 levels. On the other hand, temperature variation is the cell culture parameter which had the least effect the cell surface glycoprofile.

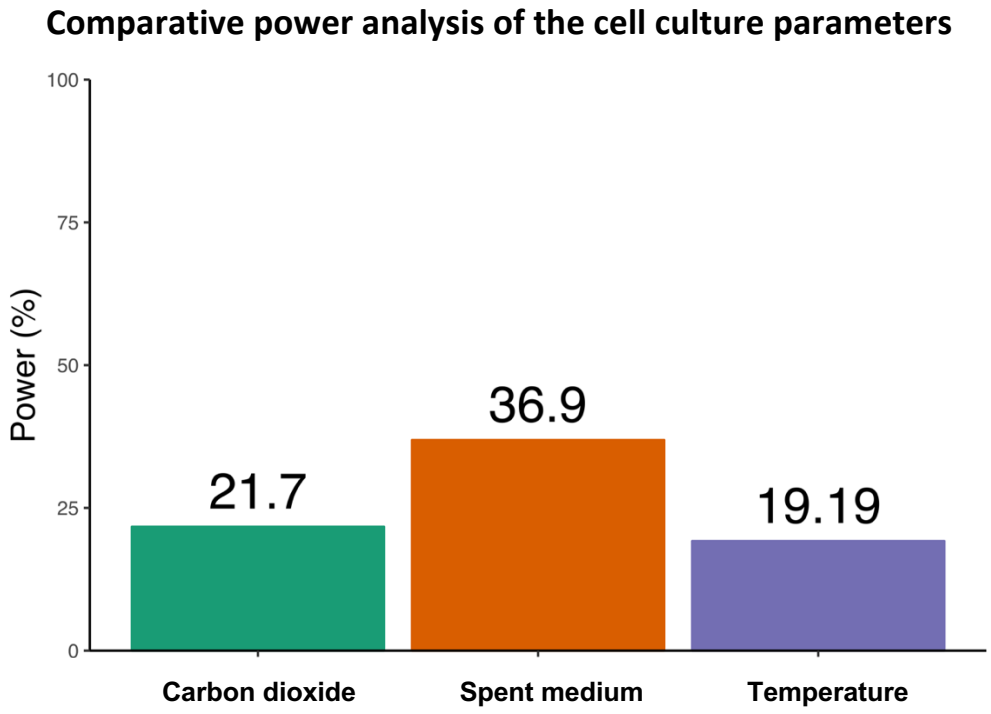

Figure 4.58: Bar plot illustrating the power averages of the values obtained upon variation of cell culture parameters. The average values were obtained by averaging the power values computed for each lectin and each parameter level (treatment).

Figure 4.59 allows the identification of the key lectin and glycan which can be scientifically meaningfully associated with alterations in the level of nutrients in the medium. As the figure shows, LEC B was the lectin with the highest power average with 57.04%. Therefore, Fucose might be a key glycan associated with the alterations in the level of spent medium. However, to a less extent, Sialic acid can be another key glycan involved in the cellular metabolism response to the variation of spent medium levels as MAL II average power was 46.05%.

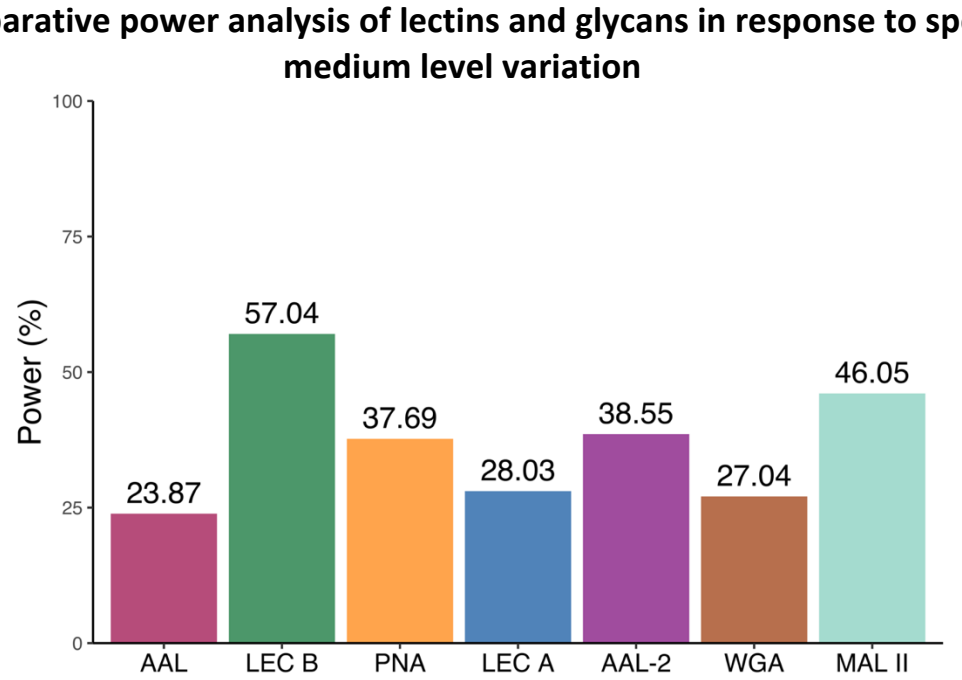

Figure 4.59: Bar plot illustrating the lectin power averages. The average values were obtained by averaging the power values computed from a lectin dataset across the 6 spent medium levels (treatments).

Whereas, N-Acetylglucosamine was identified as a key glycan associated with the variation of CO_2 and temperature levels with average power values of 33.65% and 40.12%, respectively. WGA, as can been observed in Figures 4.60 and 4.61, was the lectin which provided more scientifically meaningful results in comparison to AAL-2. In addition, Galactose and Sialic acid might be another potential key glycans which can be affected by the variation of CO_2 levels since PNA average power was 24.84% and MAL II was 24.69%. Sialic acid might also be other

potential key glycan which responds to the variation of temperature. However, MAL II average power was 19.38% which is twice as much lower than WGA average power. Table 4.7 summarizes key lectins and glycans involved in each cell culture parameter.

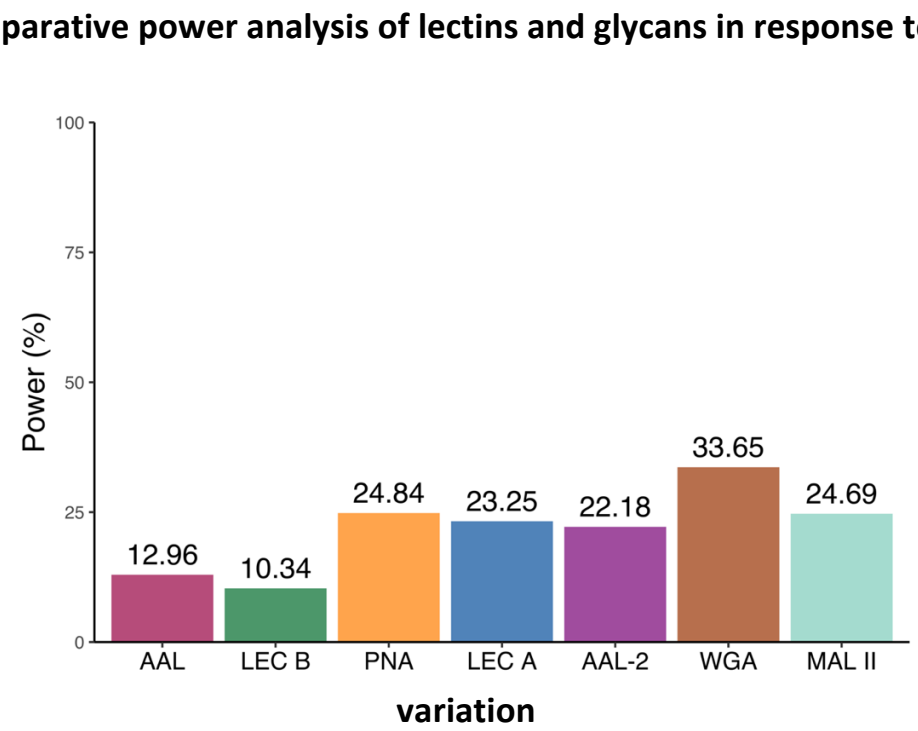

Figure 4.60: Bar plot illustrating the lectin power averages. The average values were obtained by averaging the power values computed from a lectin dataset across the 9 CO_2 levels (treatments).

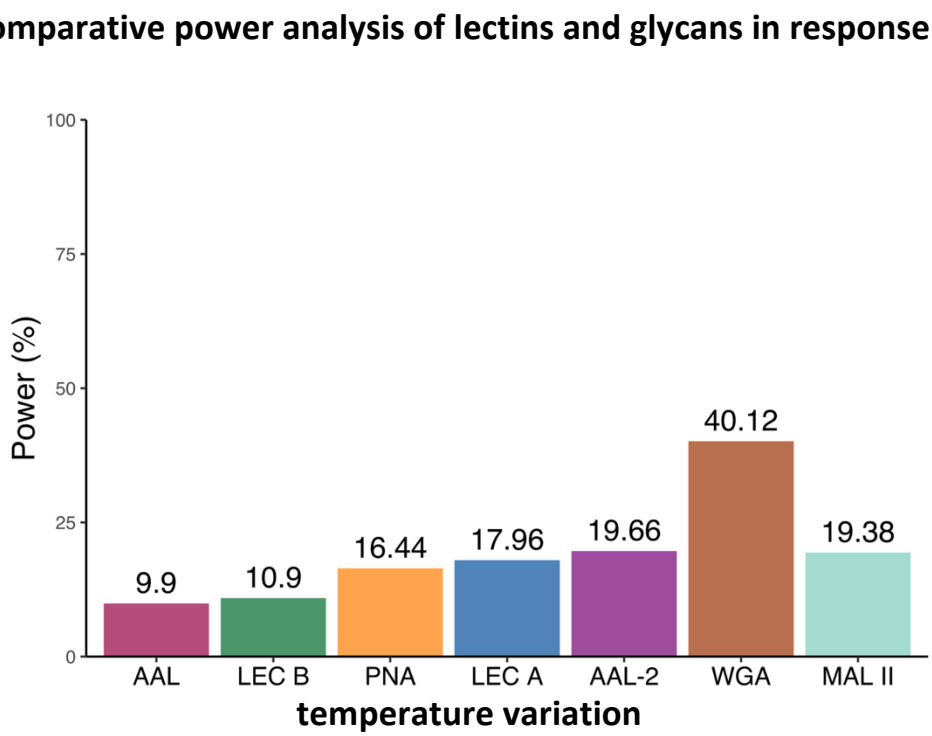

Figure 4.61: Bar plot illustrating the lectin power averages. The average values were obtained by averaging the power values computed from a lectin dataset across the 8 temperature levels (treatments).

Table 4.7: Table summarizing key lectins and glycans associated with each cell culture parameter.

Cell culture parameter	Key lectin	Key glycan	Potential key glycans
Spent medium level	LEC B	Fucose	Sialic acid
CO_2	WGA	N-Acetylglucosamine	Galactose and Sialic acid
Temperature	WGA	N-Acetylglucosamine	Sialic acid

Therefore, it can be concluded that N-Acetylglucosamine and Sialic acid are glycans with a high level of relevance in the investigation of the influence of temperature, CO_2 and nutrient levels in the cell surface glycosylation process.

4.9.6 Spent medium level variation: BCA and ELLA analysis of secreted proteins

As was concluded in the previous section, spent medium variation was identified as the cell culture parameter which most influenced the cell surface glycoprofile changes. Therefore, BCA and ELLA analysis was performed to investigate the relationship of this parameter with the concentration levels of secreted proteins and their glycoprofile. The supernatant of the cell cultures was stored at -20°C until the finalization of nutrient variation experiments. Samples were then defrosted, and their concentration levels were determined using the BCA assay (see section 3.5). Samples of 50 μL volume were then prepared at 5 μg/mL by diluting the concentrated supernatant samples with fresh supplemented medium. The diluted protein samples were then used in the ELLA assay (see section 3.9).

The results of the BCA assay can be observed in Figure 4.62 which demonstrates the relationship between the concentration levels of secreted proteins and spent medium levels. The figure shows that the concentration levels of the proteins in the medium decreased as the levels of spent medium were reduced or incremented in relation to the baseline level. However, a tendency of an increase in the concentration level can be seen between 2 and 3 days as the 4 days spent medium level is exactly the baseline level (see section 3.12.2.2).

Figure 4.63 demonstrates AAL and LEC B binding variation as the levels of spent medium were changed. It can be observed that the lectin binding pattern was the same for both lectins which specifically bind to Fucose. However, the intensity of the absorbance values is different as AAL values were higher than LEC B. The number of available Fucose sites on secreted proteins reduced as the spent medium level decreased up to -2 days, but a tendency of an increase followed by a rapid decrease in the number of Fucose took place from -2 to -3 days. Similarly, as the spent medium level increased the number of Fucose sites decreased, but it increased followed by a sharp drop between +2 and +3 days. It can be seen that the 95% confidence interval around -2 to -3 days and +2 to +3 days ranges are wider, showing a considerable increase in their level of data variability.

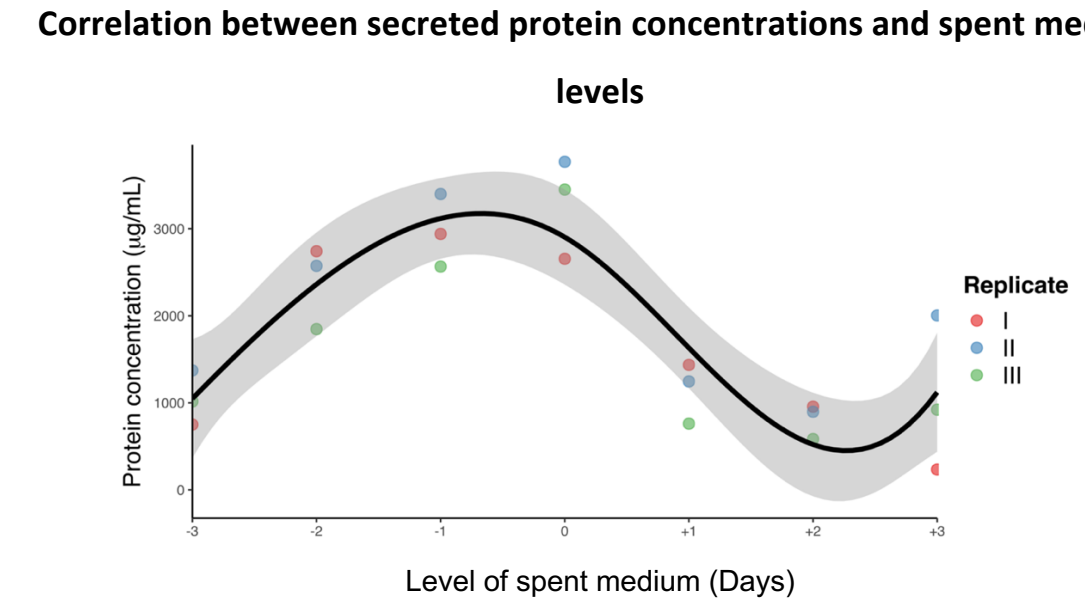

Figure 4.62: Polynomial fit with 95% confidence interval correlating the levels of secreted protein concentration and spent medium levels (see section 3.8).

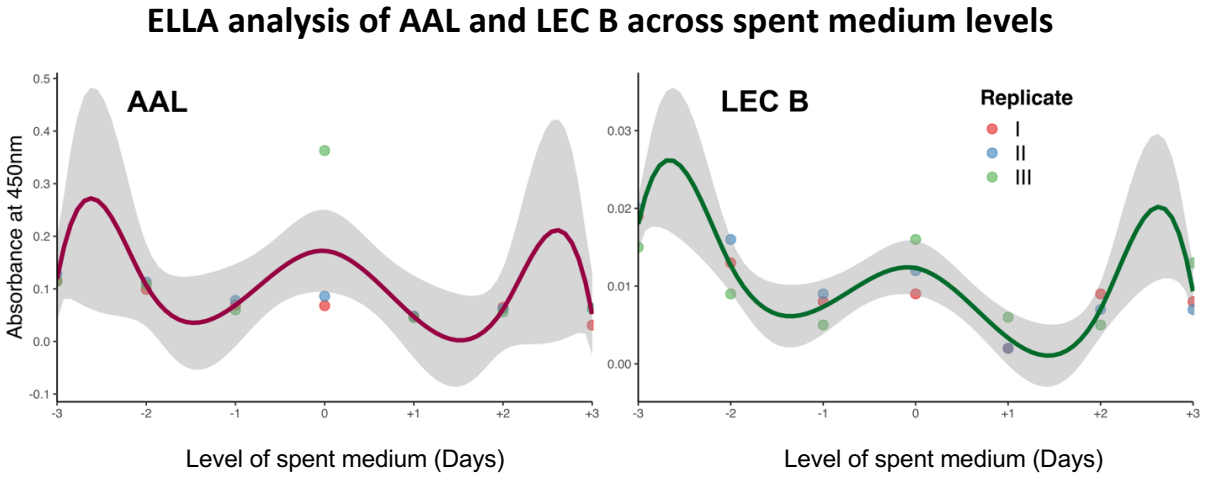

Figure 4.63: Polynomial fit with 95% confidence interval demonstrating the variation of AAL and LEC B binding as the level of spent medium changed (see section 3.9).

Figure 4.64 demonstrates PNA and LEC A binding variation as the levels of nutrients were changed. It can be observed that the lectin binding pattern was very different although both lectins specifically bind to Galactose. PNA absorbance values were around 0, while LEC A

values changed as the nutrient levels were altered. By observing LEC A polynomial curve, the number of available Galactose sites on the secreted proteins reduced in response to both an increase and decrease in the levels of nutrients. However, increased levels of nutrient caused a more significant drop in the number of Galactose. A tendency of an increase followed by a decrease in Galactose sites was observed between -2 and -3 days, and +2 and +3 days. However, the 95% confidence interval around these ranges is wider demonstrating a significant increase in their level of data variability.

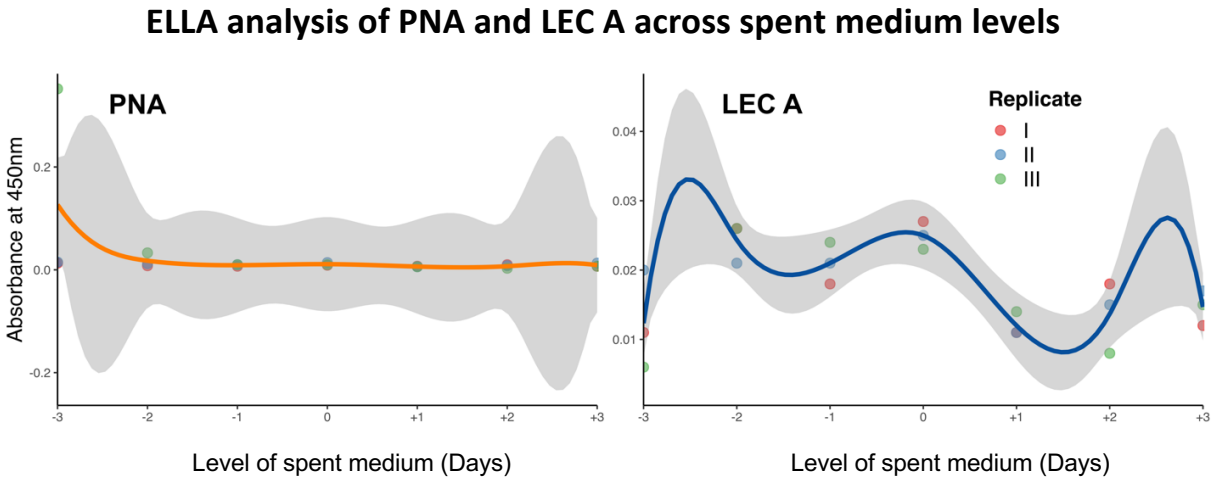

Figure 4.64: Polynomial fit with 95% confidence interval demonstrating the variation of PNA and LEC A binding as the level of spent medium changed (see section 3.9).

AAL-2 and WGA binding can be observed in Figure 4.65. Binding pattern of both lectins is the same, but WGA absorbance values were slightly higher than AAL-2. The polynomial curves showed that the number of available N-Acetylglucosamine sites on secreted proteins reduced as a result of the increase and decrease in the level of nutrients up to -2 and +2 days in relation to the baseline nutrient level. However, the curves showed a tendency of an increase followed by a rapid decrease in the lectin binding between -2 to -3 days and +2 and +3 days. It can be observed that the confidence interval around these two ranges are wider showing a considerable increase in their level of data variability.

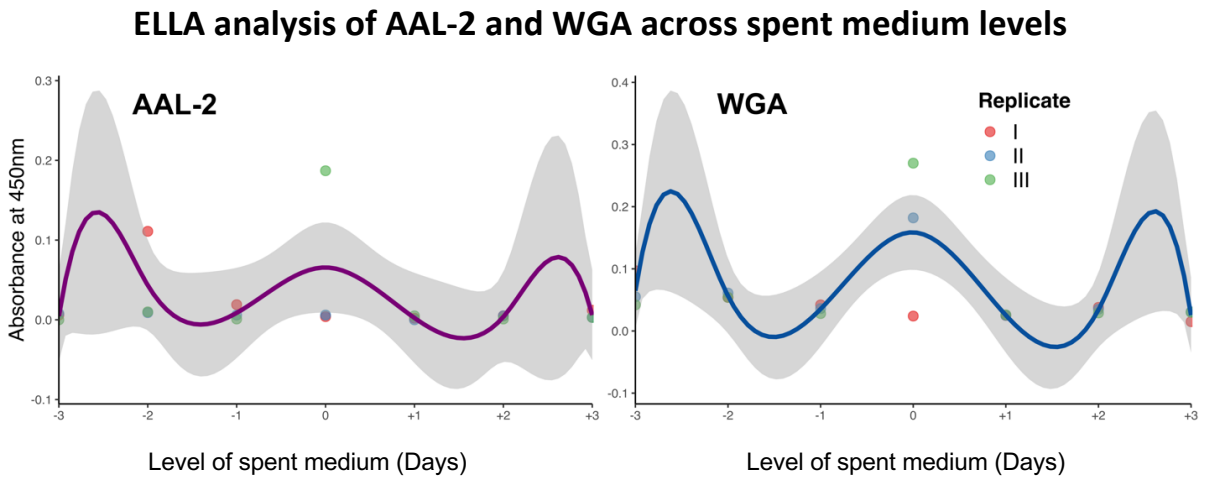

Figure 4.65: Polynomial fit with 95% confidence interval demonstrating the variation of AAL-2 and WGA binding as the level of spent medium changed (see section 3.9).

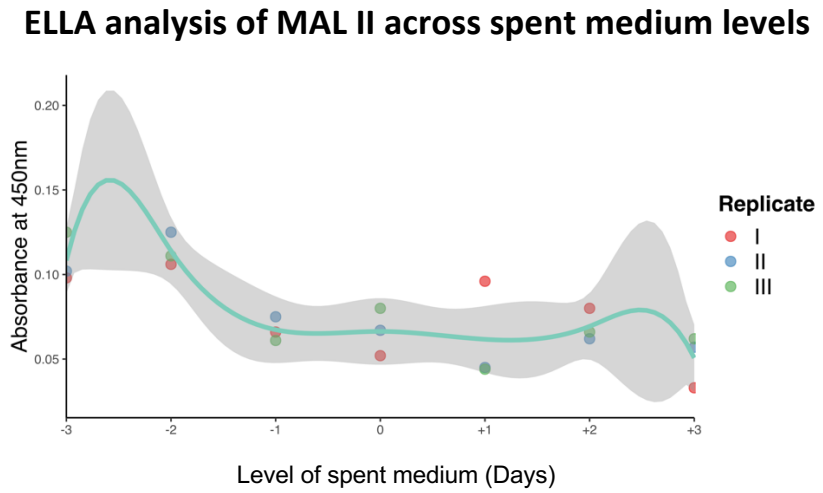

Figure 4.66: Polynomial fit with 95% confidence interval demonstrating the variation of MAL II binding as the level of spent medium changed (see section 3.9).

The variation in the number of available Sialic acid sites on secreted proteins as the level of nutrients was changed can be observed in Figure 4.66. The number of Sialic acid sites was constant up to -1 and +2 days in relation to the baseline nutrient level. However, a considerable increase can be seen as a result of further depleted levels of nutrients. A tendency of a rapid decrease in the number of glycan sites was shown between -2 and -3 days.

Similarly, a decrease tendency in the Sialic acid sites was demonstrated between +2 and +3 days. However, the level of data variability increased around these two ranges since the 95% confidence interval is wider.

The BCA and ELLA results on secreted proteins demonstrated that the variation of nutrient levels in the medium has an effect on the concentration levels of secreted proteins and their glycoprofile. Although the ELLA assay is not as quantitative as flow cytometry to allow for a meaningful indication of the direction of protein glycoprofile changes, the results presented in this section have shown that the glycoprofile of secreted proteins can also be altered due to the variation of nutrient levels in the medium. Moreover, it can be concluded that the cell surface glycoprofile alterations associated with the variation of nutrient levels can be a strong indication of glycoprofile changes in secreted proteins. Therefore, the flow cytometric methodology here presented can be very useful as a monitoring tool allowing the assessment of the health status of CHO-K1 cells, enhancing the ability of detecting early alterations in cellular metabolism by probing the cell surface with appropriate lectins.

4.9.7 Statistical analysis of lectin interaction with cell surface throughout the DNA cell cycle

As was observed (see sections 4.9.1, 4.9.2 and 4.9.3), the intensity of LECTIN-A signal increased across the DNA cell cycle. In other words, LECTIN-A signal from G2/M cells was the strongest, while the signal from Go/G1 cells was the weakest (Go/G1 < S < G2/M). Since FSC-A data of the DNA subpopulations also increased as cells went from Go/G1 to G2/M cell cycle stages, it has been concluded that the increase in LECTIN-A signal could be due to the enlargement of the cells towards cell replication. Another observation which further supported this conclusion was the repetition of LECTIN-A Go/G1 curve pattern in S and G2/M curves.

In order to exclude the cell size as the factor which could influence the level of lectin interaction on cell surface, each cell LECTIN-A value was divided by its FSC-A value (a relative

cell size parameter). Thus, the level of lectin interaction can be interpreted as the lectin density on cell surface which, in turn, can be directly compared across the DNA cell cycle subpopulations. Therefore, this section aims to analyse lectin density data for the identification of a pattern different than Go/G1 < S < G2/M (see section 3.12.7.4).

Figure 4.67 demonstrates the lectin density values across the DNA cell cycle under the nutrient levels which cells were subjected to. The most common pattern is Go/G1 < S < G2/M showing that the number of glycans available to interact with lectins increased as the cells went through the DNA cell cycle. In other words, glycans may have been upregulated through the upregulation of proteins involved with the DNA cell cycle. For instance, a study with HeLa cells used a quantitative proteomic approach to compare cell surface-exposed proteins in mitosis and interphase. Out of the 628 surface and surface associated proteins identified in HeLa cells, 27 were considerably enriched at the cell surface in mitosis and 37 in interphase. The proteins which were regulated by the cell cycle were involved in cell adhesion, receptor, and endosome/lysosome biology. However, it was found that adhesion biomolecules were one of the most prominent classes of proteins whose cell surface exposure altered during the progression of mitosis (Ozlu *et al.*, 2015).

Lectin interaction with cell surface throughout the DNA cell cycle in response to spent medium level variation

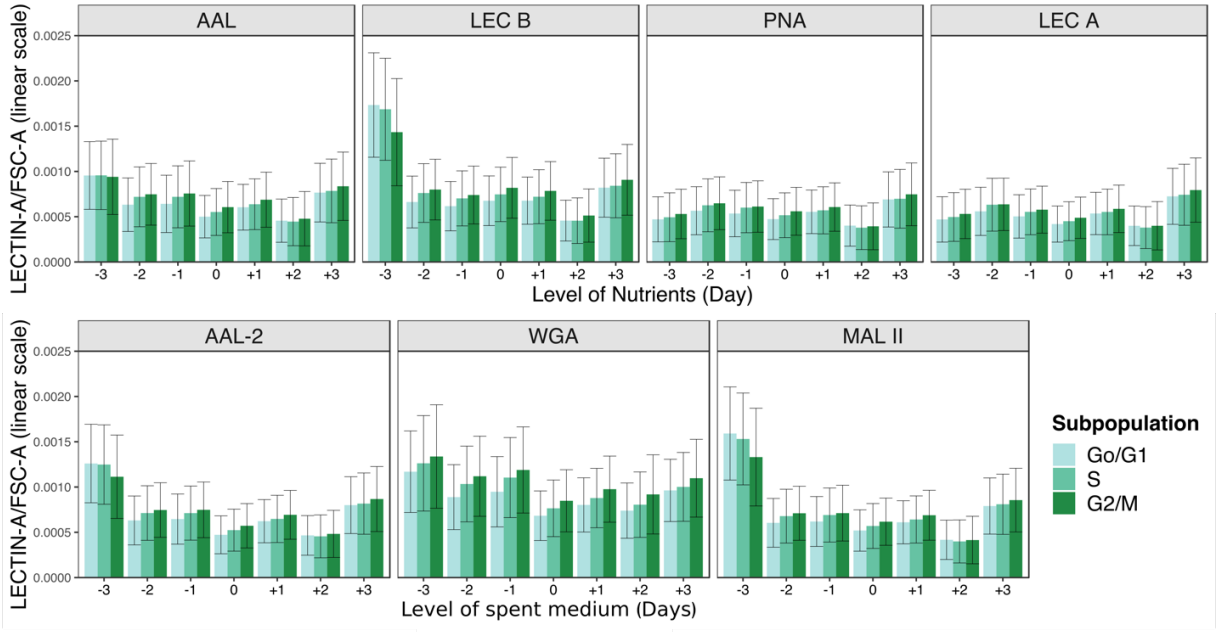

Figure 4.67: Bar plot demonstrating the alterations of lectin interaction with cell surface as cells go through the DNA cell cycle under different levels of spent medium. Data obtained through a BD FACSAria™ I flow cytometer using the ratio of LECTIN-A and FSC-A signals of each interrogated cell (see section 3.12.7.4).

Figure 4.67 also reveals a second pattern in which glycans seemed to have been downregulated from Go/G1 to G2/M, that is, G2/M < S < Go/G1. This pattern was found at -3 day of LEC B, AAL-2, MAL II; therefore, Fucose, N-Acetylglucosamine and Sialic acid glycans, respectively. A third pattern can be observed, in which Fucose, Galactose, N-Acetylglucosamine and Sialic acids may have been downregulated from Go/G1 to S, then the glycans were upregulated from S to G2/M to the same levels at Go/G1. This pattern was identified at +2 day, AAL, PNA, LEC A, AAL-2, and MAL II. However, at +2 day of LEC B, this glycan regulation pattern was not clear.

In the case of cells subjected to the variation of temperature levels, Figure 4.68 demonstrates the lectin density values across the DNA cell cycle under the temperature levels. As was also

observed in Figure 4.67, the most common pattern is Go/G1 < S < G2/M showing that the number of glycans available to interact with lectins increased as the cells went through the DNA cell cycle. However, the figure reveals a second pattern in which glycans seemed to have been downregulated from Go/G1 to G2/M, that is, G2/M < S < Go/G1. This pattern was found at +3 °C of AAL, LEC B, and MAL II; therefore, Fucose and Sialic acid glycans, respectively.

Lectin interaction with cell surface throughout the DNA cell cycle in response to temperature variation

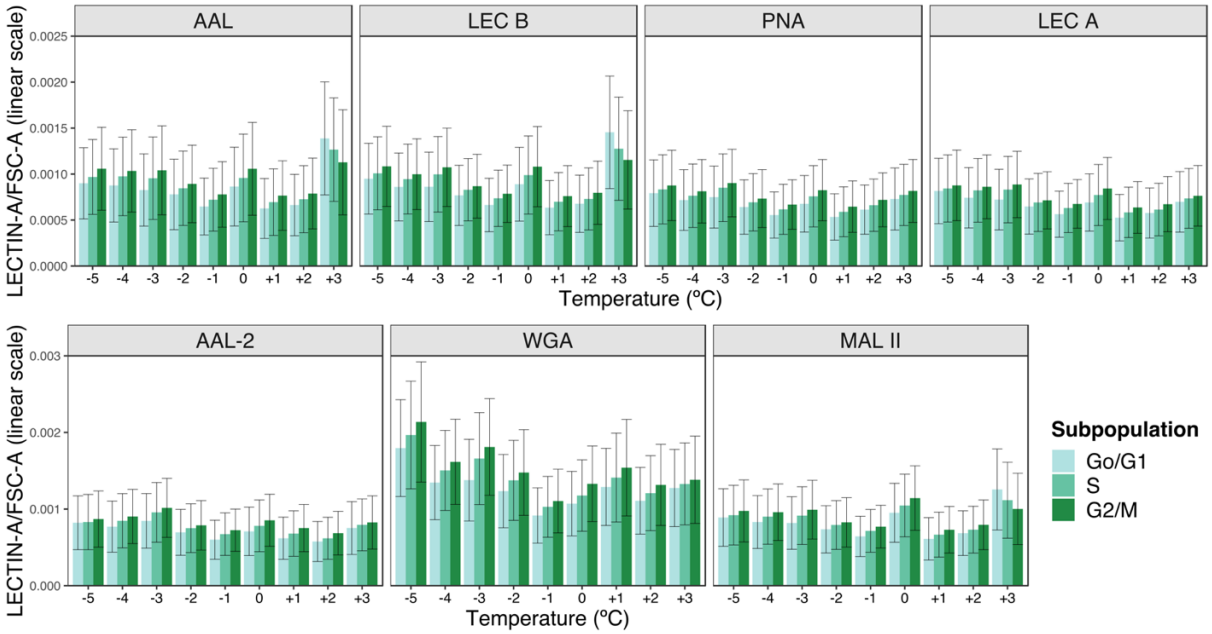

Figure 4.68: Bar plot demonstrating the alterations of lectin interaction with cell surface as cells go through the DNA cell cycle under different levels of temperature. Data obtained through a BD FACSAria™ I flow cytometer using the ratio of LECTIN-A and FSC-A signals of each interrogated cell (see section 3.12.7.4).

Figure 4.69 demonstrates the lectin density values across the DNA cell cycle under the CO_2 levels which cells were subjected to. As was also observed in Figures 4.67 and 4.68, the most common pattern is Go/G1 < S < G2/M showing that the number of glycans available to interact with lectins increased as the cells went through the DNA cell cycle. However, the figure reveals a second pattern at +1 % of AAL, PNA, and AAL-2, in which Fucose, Galactose and N-

Acetylglucosamine, respectively, may have been upregulated from Go/G1 to S, then downregulated to a higher level in relation to Go/G1 from S to G2/M cell cycle phases. The most apparent pattern is then Go/G1 < G2/M < S, but this pattern is not clear from PNA data.

Lectin interaction with cell surface throughout the DNA cell cycle in response to CO_2 variation

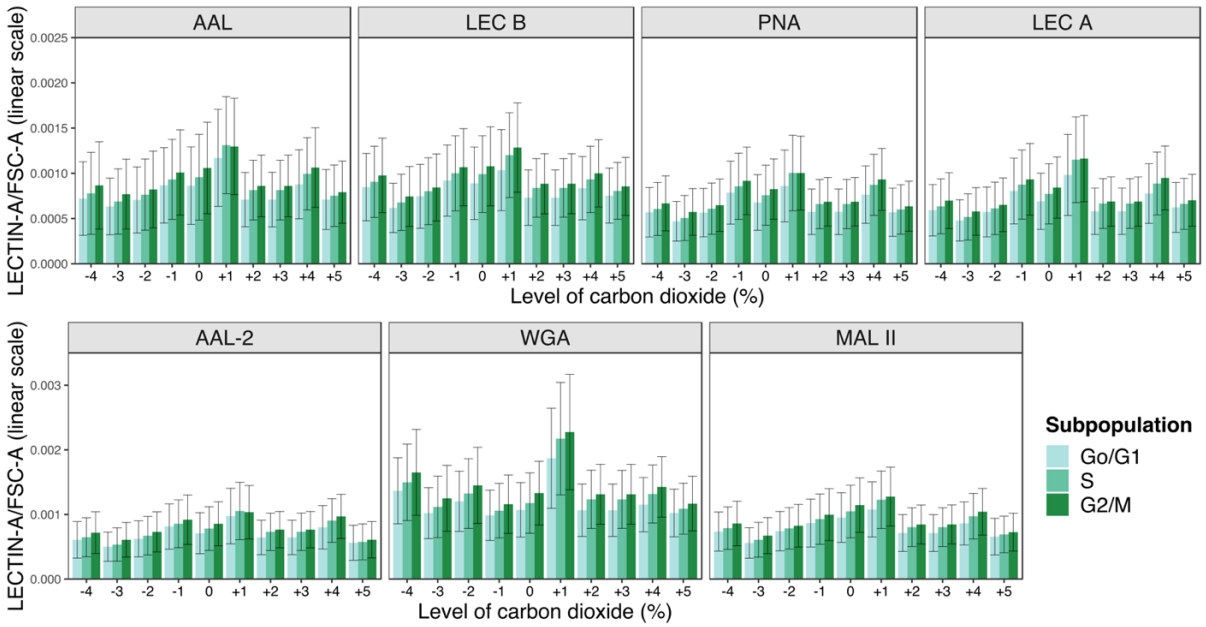

Figure 4.69: Bar plot demonstrating the alterations of lectin interaction with cell surface as cells go through the DNA cell cycle under different levels of CO_2. Data obtained through a BD FACSAria™ I flow cytometer using the ratio of LECTIN-A and FSC-A signals of each interrogated cell (see section 3.12.7.4).

The fact that the pattern Go/G1 < S < G2/M was the most common one found in the experiments can further support the conclusion that the number of cell surface glycans available to interact with lectins increased as the cells went through the DNA cell cycle, and such increase may not be associated with cell enlargement which occurs towards cell mitosis. In fact, the increase is likely to be associated with cellular metabolism modifications related to the DNA cell cycle. For instance, a number of scientific studies has reported the changes in the levels of O-GlcNAc throughout the cell cycle (Lefebvre *et al.*, 2004; Slawson *et al.*, 2005;

Drougat *et al.*, 2012; Fong *et al.*, 2012). Furthermore, the changes in Go/G1 < S < G2/M pattern of some glycans in response to certain levels of nutrient, temperature and CO_2, have demonstrated that these cell culture parameters can also influence the normal cell surface glycoprofile alterations throughout the cell cycle. Most importantly, that influence was commonly identified in parameter levels which eventually led cell viability into a dramatic decrease (Figures 4.31, 4.40 and 4.49), particularly concerning to the variation of spent medium levels, as was also concluded in section 4.9.4. Therefore, the alterations in the normal cell surface glycoprofile changes across the DNA cell cycle could be an indicative of cellular metabolism modifications associated with harmful growing conditions.

5 Final considerations and future work

The methodology here described has demonstrated that flow cytometry may be used to investigate the variation in cell surface glycosylation using lectins as probes since scientifically meaningful data was obtained (see sections 4.9.1-3). Furthermore, the methodology demonstrated that early meaningful changes in cell surface glycosylation can be detected and the use of DNA and viability dyes (DRAQ5 and 7-AAD) is important as cell surface glycosylation is associated with the DNA cycle and dramatic changes can be seen under very stressful growing conditions (see section 4.9.6).

However, the methodology provided scientifically meaningful data from Go/G1 subpopulation only. Since the glycosylation on the surface of a cell is associated with its DNA cycle, it is relevant to further develop the methodology to investigate cell surface glycosylation of G2/M and S subpopulations. This may be achieved by increasing the number of interrogated cells. However, the length of time required to collect the data would also increase.

Although the methodology has looked at five different glycans on the cell surface (see section 4.3), the expansion of the lectin panel would allow a more comprehensive analysis of the cell surface glycosylation changes thus increasing the range of glycan types investigated. This would also mean a concurrent and significant increase in both the complexity and size of the data. Therefore, more advanced techniques to analyze the data might be of more relevance such as machine learning and perhaps deep learning techniques. In fact, these techniques could have been applied and demonstrated in this PhD thesis since the dataset is sufficiently large and complex. However, classical statistical techniques were the choice for the thesis while machine learning techniques will be applied to the dataset to be published in the form of an article paper in the near future. An interesting outcome of this future work is the development of a predictive model correlating cell surface glycosylation changes with the cell culture process parameters that were altered.

The research work here presents an analytical system that has the potential to be utilized in the bioprocess monitoring of cells during the bioreactor step. Since cell surface glycosylation

of a bioprocessing cell is usually associated with the glycosylation of secreted proteins (see section 4.9.5), monitoring the cell surface may be a powerful alternative to ensure that the glycosylation of therapeutic proteins, a critical quality attribute, is within the desired parameters. The methodology offers a rapid and automated manner to evaluate the glycosylation on cell surface in relation to a healthy cell surface glycoprofile. Once the samples have been prepared for flow cytometric analysis and data have been collected, the methodology offers an automated analysis generating multiple formats to visualize the data. The entire process can be concluded within 5 to 6 hours.

In addition, although these studies were conducted on CHO-K1 cells, the methodology may be applied to other cell lines. Therefore, this research work can be the foundation of a more sophisticated methodology to interrogate human cells in order to aid the diagnosis of diseases.

6 Conclusion

In conclusion, CHO-K1 cell surface glycosylation alters in response to spent medium, temperature and CO_2 variation. However, the changes observed are most scientifically meaningful when looking at cell surface glycosylation variation in reponse to spent medium levels. In addition, fucosylated glycoforms may be a key carbohydrate structure changing on cell surface in response to spent medium levels, while GlcNAcylated glycoforms are associated with temperature and CO_2 alteration.

The combination of the use of lectin probes and the flow cytometric methodology which was developed allows the early detection of changes in cell surface glycosylation. This methodology also provides information on cell surface glycosylation alteration as cells go through the DNA cycle. Consequently, the methodology may be used to monitor the bioprocessing cell in order to detect early changes on cell surface glycosylation associated with stressful growing conditions. This could allow for timely remedial action to be taken that could potentially save the entire production batch. Therefore, the implementation of such technology in an industrial setting would greatly increase the knowledge of the bioprocess and the ability to monitor and control it, ensuring the quality of the protein of interest.

7 References

Adan, A. *et al.* (2017) 'Flow cytometry: basic principles and applications.', *Critical reviews in biotechnology*, 37(2), pp. 163–176. doi: 10.3109/07388551.2015.1128876.

Aghamohseni, H. *et al.* (2014) 'Effects of nutrient levels and average culture pH on the glycosylation pattern of camelid-humanized monoclonal antibody', *Journal of Biotechnology*. Elsevier B.V., 186, pp. 98–109. doi: 10.1016/j.jbiotec.2014.05.024.

Akama, T. O. *et al.* (2002) 'Germ Cell Survival Through Interaction with Sertoli Cells', *Science*, 295(5552), pp. 124–127.

Al-Fageeh, M. B. and Smales, C. M. (2006) 'Control and regulation of the cellular responses to cold shock: the responses in yeast and mammalian systems', *Biochemical Journal*, 397(2), pp. 247–259. doi: 10.1042/BJ20060166.

Amano, J., Kobayashi, K. and Oshima, M. (2001) 'Comparative Study of Glycosyltransferase Activities in Caco-2 Cells before and after Enterocytic Differentiation Using Lectin-Affinity High-Performance', *Archives of Biochemistry and Biophysics*, 395(2), pp. 191–198. doi: 10.1006/abbi.2001.2572.

An, H. J. *et al.* (2009) 'Glycomics and disease markers', *Current Opinion in Chemical Biology*, 13(5–6), pp. 601–607. doi: 10.1016/j.cbpa.2009.08.015.

Baik, J. Y. *et al.* (2006) 'Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin', *Biotechnology and Bioengineering*, 93(2), pp. 361–371. doi: 10.1002/bit.20717.

Balcan, E. *et al.* (2008) 'Cell surface glycosylation diversity of embryonic thymic tissues', *Acta histochemica*, 110(1), pp. 14–25. doi: 10.1016/j.acthis.2007.07.003.

Bandaranayake, A. D. and Almo, S. C. (2014) 'Recent advances in mammalian protein production', *FEBS Letters*. 588(2), pp. 253–260. doi: 10.1016/j.febslet.2013.11.035.

Bates, M. K., Phillips, D. S. and O'Bryan, J. (2011) 'Shaker Agitation Rate and Orbit Affect

Growth of Cultured Bacteria', *Thermo Scientific*. Available at: www.thermoscientific.com/. Acessed on 10/02/2018.

Batisse, C. *et al.* (2004) 'Lectin-based three-color flow cytometric approach for studying cell surface glycosylation changes that occur during apoptosis', *Cytometry Part A*, 62(2), pp. 81–88. doi: 10.1002/cyto.a.20094.

Bierhuizen, M. F. A. and Fukuda, M. (1992) 'Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6GlcNAc transferase by gene transfer into CHO cells expressing polyoma large tumor antigen', *Proceedings of the National Academy of Sciences of the United States of America*, 89(October), pp. 9326–9330.

Biosciences, B. (2009) 'An introduction to compensation for multicolor assays on digital flow cytometers', *BD Biosciences Technical Bulletin*, (August), pp. 1–12.

Bollati-Fogolín, M. *et al.* (2005) 'Temperature Reduction in Cultures of hGM-CSF-expressing CHO Cells: Effect on Productivity and Product Quality - Bollati-Fogolín - 2008 - Biotechnology Progress - Wiley Online Library', *Biotechnology progress*, 21(1), pp. 17–21. doi: 10.1021/bp049825t.

Boron, W. F. (1987) 'Control of intracellular pH', in Brenner, B. M. and Stein, J. H. (eds) *Modern Techniques of ion transport*. New York: Churchill Livingstone.

Borys, M. C., Linzer, D. I. H. and Papoutsakis, E. T. (1994) 'Ammonia affects the glycosylation patterns of recombinant mouse placental lactogen-I by chinese hamster ovary cells in a pH-dependent manner', *Biotechnology and Bioengineering*, 43(6), pp. 505–514. doi: 10.1002/bit.260430611.

Boscher, C., Dennis, J. W. and Nabi, I. R. (2011) 'Glycosylation, galectins and cellular signaling', *Current Opinion in Cell Biology*. Elsevier Ltd, 23(4), pp. 383–392. doi: 10.1016/j.ceb.2011.05.001.

Brandon, M., Baldi, P. and Wallace, D. C. (2006) 'Mitochondrial mutations in cancer', *Oncogene*, 25(3), pp. 4647–4662. doi: 10.1385/EP:17:3:203.

van Breukelen, F. and L. Martin, S. (2002) 'Molecular adaptations in mammalian hibernators: unique adaptations or generalized responses?', *J Appl Physiol*, 92, pp. 2640–2647.

Brockhausen, I., Romero, P. A. and Herscovics, A. (1991) 'Glycosyltransferase Changes upon Differentiation of CaCo-2 Human Colonie Adenocarcinoma Cells', *Cancer Research*, 51(12), pp. 3136–3142.

Brockhausen, I., Vavasseur, F. and Yang, X. (2001) 'Biosynthesis of mucin type O-glycans: Lack of correlation between glycosyltransferase and sulfotransferase activities and CFTR expression', *Glycoconjugate Journal*, 18(9), pp. 685–697.

Broger, T. *et al.* (2011) 'Real-time on-line flow cytometry for bioprocess monitoring', *Journal of Biotechnology*. Elsevier B.V., 154(4), pp. 240–247. doi: 10.1016/j.jbiotec.2011.05.003.

Brooks, S. A. (2017) 'Lectin Histochemistry: Historical Perspectives, State of the Art, and the Future', *Methods in Molecular Biology*, 1560, pp. 93–107. doi: 10.1007/978-1-4939-6788-9.

Campbell, C. and Stanley, P. (1984) 'A dominant mutation to ricin resistance in Chinese hamster ovary cells induces UDP-GlcNAc: glycopeptide beta-4-N-acetylglucosaminyltransferase III activity', *The Journal of Biological Chemistry*, 261(21), pp. 13370–13378.

Chakrabarti, A., Chen, A. W. and Varner, J. D. (2011) 'A Review of the Mammalian Unfolded Protein Response', *Biotechnology and Bioengineering*, 108(12), pp. 2777–2793. doi: 10.1002/bit.23282.

Chen, S. *et al.* (2007) 'Analysis of cell surface carbohydrate expression patterns in normal and tumorigenic human breast cell lines using lectin arrays', *Analytical Chemistry*, 79(15), pp. 5698–5702. doi: 10.1021/ac070423k.

Chen, X. *et al.* (2010) 'Quantitative proteomics analysis of cell cycle-regulated golgi disassembly and reassembly', *Journal of Biological Chemistry*, 285(10), pp. 7197–7207. doi: 10.1074/jbc.M109.047084.

Chrispeels, M. J. and Raikhelb, N. V (1991) 'Lectins, Lectin Genes, and Their Role in Plant

Defense', *The Plant Cell*, 3(1), pp. 1–9.

Christiansen, M. N. *et al.* (2014) 'Cell surface protein glycosylation in cancer', *Proteomics*, 14(4–5), pp. 525–546. doi: 10.1002/pmic.201300387.

Clark, K. J. R., Chaplin, F. W. R. and Harcum, S. W. (2004) 'Temperature effects on product-quality-related enzymes in batch CHO cell cultures producing recombinant tPA', *Biotechnology Progress*, 20(6), pp. 1888–1892. doi: 10.1021/bp049951x.

Conrad, V. K. *et al.* (2019) 'Implementation and Validation of an Automated Flow Cytometry Analysis Pipeline for Human Immune Profiling', *Cytometry Part A*, 95(2), pp. 183–191. doi: 10.1002/cyto.a.23664.

Van Damme, E. J. M., Lannoo, N. and Peumans, W. J. (2008) 'Plant Lectins', *Advances in Botanical Research Incorporating Advances in Plant Pathology*, 48, pp. 107–209. doi: 10.1016/S0065-2296(08)00403-5.

Delannoy, P. *et al.* (2017) 'Glycosylation Changes Triggered by the Differentiation of Monocytic THP - 1 Cell Line into Macrophages', *Journal of Proteome Research*, 16(1), pp. 156–169. doi: 10.1021/acs.jproteome.6b00161.

Demain, A. L. and Vaishnav, P. (2009) 'Production of recombinant proteins by microbes and higher organisms', *Biotechnology Advances*. Elsevier Inc., 27(3), pp. 297–306. doi: 10.1016/j.biotechadv.2009.01.008.

Dennis, J. M. (1992) 'Changes in glycosylation associated with malignant transformation and tumor progression', in Fukuda, M. (ed.) *Cell Surface Carbohydrates and Cell Development*. Boca Raton: CRC Press, pp. 161–194.

Dennis, J. W. *et al.* (1989) 'Oncogenes conferring metastatic potential induce increased branching of Asn-linked oligosaccharides in rat2 fibroblasts', *Oncogene*, 4(7), pp. 853–860.

Dennis, J. W. (1991) 'N-linked oligosaccharide processing and tumor cell biology', *Seminar in Cancer Biology*, 2(6), pp. 411–20.

Dennis, J. W., Lau, K. S. and Nabi, I. R. (2009a) 'Adaptive Regulation at the Cell Surface by N-

Glycosylation', *Traffic*, 10(11), pp. 1569–1578. doi: 10.1111/j.1600-0854.2009.00981.x.

Dennis, J. W., Nabi, I. R. and Demetriou, M. (2009b) 'Metabolism, Cell Surface Organization, and Disease', *Cell*, 139(7), pp. 1229–1241. doi: 10.1016/j.cell.2009.12.008.

Deschenes, I., Finkle, C. and Winocour, P. D. (1997) 'Effective use of BCH-2763, a new potent injectable direct thrombin inhibitor, in combination with tissue plasminogen activator (TPA) in a rat arterial thrombolysis model', *FASEB Journal*, 11(3), pp. 186–191.

Dietmair, S. *et al.* (2012) 'A multi-omics analysis of recombinant protein production in Hek293 cells', *PLoS ONE*, 7(8). doi: 10.1371/journal.pone.0043394.

Drougat, L. *et al.* (2012) 'Characterization of O-GlcNAc cycling and proteomic identification of differentially O-GlcNAcylated proteins during G1/S transition', *Biochimica et Biophysica Acta - General Subjects*. Elsevier B.V., 1820(12), pp. 1839–1848. doi: 10.1016/j.bbagen.2012.08.024.

Esko, J. D., Stewart, T. E. and Taylor, W. H. (1985) 'Animal cell mutants defective in glycosaminoglycan biosynthesis', *Proceedings of the National Academy of Sciences of the United States of America*, 82(10), pp. 3197–3201.

Feizi, T. and Mulloy, B. (2003) 'Carbohydrates and glycoconjugates Glycomics : the new era of carbohydrate biology', *Current Opinion in Structural Biology*, 13(5), pp. 602–604. doi: 10.1016/j.sbi.2003.09.001.

Fernandes, B. *et al.* (1991) 'Beta 1-6 Branched Oligosaccharides as a Marker of Tumor Progression in Human Breast and Colon Neoplasia', *Cancer Research*, 51(2), pp. 718–723.

Ferrini, J.-B. *et al.* (1995) 'Expression of functional ricin B chain using the baculovirus system', *European Journal of Biochemistry*, 233(3), pp. 772–777.

Fong, J. J. *et al.* (2012) 'β- N -Acetylglucosamine (O -GlcNAc) Is a Novel Regulator of Mitosis-specific Phosphorylations on Histone H3', *Journal of Biological Chemistry*, 287(15), pp. 12195–12203. doi: 10.1074/jbc.m111.315804.

Forcinio, H. (2003) 'Pharmaceutical Industry Embraces NIR Technology', *Spectroscopy*, 18(9), pp. 16–24.

Fouquaert, E. *et al.* (2009) 'Related lectins from snowdrop and maize differ in their carbohydrate-binding specificity', *Biochemical and Biophysical Research Communications*. Elsevier Inc., 380(2), pp. 260–265. doi: 10.1016/j.bbrc.2009.01.048.

Frankel, A. *et al.* (1994) 'Expression of ricin B chain in *Spodoptera frugiperda*', *Biochemical Journal*, 303(Pt 3), pp. 787–794.

Freeze, H.H. and Schachter, H. (2009) 'Genetic disorders of glycosylation', in Varki, A. et al. (eds) *Essentials of glycobiology*. 2nd edn. New York: Cold Spring Harbor Laboratory Press, pp. 585-600.

Gemeiner, P. *et al.* (2009) 'Lectinomics II. A highway to biomedical/clinical diagnostics', *Biotechnology Advances*, 27(1), pp. 1–15. doi: 10.1016/j.biotechadv.2008.07.003.

Ghazarian, H., Idoni, B. and Oppenheimer, S. B. (2011) 'A glycobiology review : Carbohydrates, lectins and implications in cancer therapeutics', *Acta histochemica*. Elsevier, 113(3), pp. 236–247. doi: 10.1016/j.acthis.2010.02.004.

Giomarelli, B. *et al.* (2006) 'Recombinant production of anti-HIV protein, griffithsin, by auto-induction in a fermentor culture', *Protein Expression and Purification*, 47(1), pp. 194–202. doi: 10.1016/j.pep.2005.10.014.

Gorelik, E., Galili, U. and Raz, A. (2001) 'On the role of cell surface carbohydrates and their binding proteins (lectins) in tumor metastasis', *Cancer and Metastasis Reviews*, 20(3), pp. 245–277.

Grainger, R. K. and James, D. C. (2013) 'CHO Cell Line Specific Prediction and Control of Recombinant Monoclonal Antibody N-Glycosylation', *Biotechnology and Bioengineering*, 110(11), pp. 2970–2983. doi: 10.1002/bit.24959.

Graves, S. W. and Pearlson, A. (2013) 'Fluidics', *Current Protocols in Cytometry*, 29(6), pp. 997–1003. doi: 10.1016/j.biotechadv.2011.08.021.Secreted.

Gupta, G., Surolia, A. and Sampathkumar, S. (2010) 'Lectin Microarrays for Glycomic Analysis', *OMICS A Journal of Integrative Biology*, 14(4), pp. 419–436. doi: 10.1089/omi.2009.0150.

Halliwell, B. (2007) 'Oxidative stress and cancer: have we moved forward?', *Biochemical Journal*, 401(1), pp. 1–11. doi: 10.1042/bj20061131.

Hamanaka, R. B. and Chandel, N. S. (2010) 'Mitochondrial reactive oxygen species regulate cellular signaling and dictate biological outcomes', *Trends in Biochemical Sciences*. Elsevier Ltd, 35(9), pp. 505–513. doi: 10.1016/j.tibs.2010.04.002.

Hamouda, H. *et al.* (2014) 'Rapid analysis of cell surface N-glycosylation from living cells using mass spectrometry', *Journal of Proteome Research*, 13(12), pp. 6144–6151. doi: 10.1021/pr5003005.

Hanover, J. A. (2001) 'Glycan-dependent signaling : O-linked N-acetylglucosamine', *The FASEB Journal*, 15(11), pp. 1865–1876.

Harold, R. and Gabius, H.-J. (2001) 'Plant lectins: Occurrence, biochemistry, functions and applications', *Glyconjugate Journal*, 18(8), pp. 589–613.

Hart, G. W., Housley, M. P. and Slawson, C. (2007) 'Cycling of O-linked β - N -acetylglucosamine on nucleocytoplasmic proteins', *Nature*, 446(26 April), pp. 1017–1022. doi: 10.1038/nature05815.

Hirabayashi, J. (2008) 'Concept, strategy and realization of lectin-based glycan profiling', *Journal of Biochemistry*, 144(2), pp. 139–147. doi: 10.1093/jb/mvn043.

De Hoff, P., Brill, L. M. and Hirsch, A. M. (2009) 'Plant lectins: the ties that bind in root symbiosis and plant defense', *Molecular Genetics and Genomics*, 282(1), pp. 1–15. doi: 10.1007/s00438-009-0460-8.

Hong, Y. *et al.* (2004) 'The Lec23 Chinese hamster ovary mutant is a sensitive host for detecting mutations in alpha-glucosidase I that give rise to congenital disorder of glycosylation IIb (CDG IIb)', *The Journal of Biological Chemistry*, 279(48), pp. 49894–49901. doi: 10.1074/jbc.M410121200.

Howard, D. R. *et al.* (1987) 'The GDP-fucose: N-acetylglucosaminide 3-alpha-L-fucosyltransferases of LEC11 and LEC12 Chinese hamster ovary mutants exhibit novel

specificities for glycolipid substrates', *The Journal of Biological Chemistry*, 262(35), pp. 16830–16837.

Hsu, K.-L., Gildersleeve, J. C. and Mahal, L. K. (2008) 'A simple strategy for the creation of a recombinant lectin microarray', *Molecular BioSystems*, 4(6), pp. 654–662. doi: 10.1039/b800725j.

Jadhav, V. *et al.* (2013) 'CHO microRNA engineering is growing up: Recent successes and future challenges', *Biotechnology Advances*, 31(8), pp. 1501–1513. doi: 10.1016/j.biotechadv.2013.07.007.

Jagtap, U. B. and Bapat, V. A. (2010) 'Artocarpus: A review of its traditional uses, phytochemistry and pharmacology', *Journal of Ethnopharmacology*. Elsevier Ireland Ltd, 129(2), pp. 142–166. doi: 10.1016/j.jep.2010.03.031.

Jayapal, K. P. *et al.* (2007) 'Recombinant Protein Therapeutics from CHO Cells — 20 Years and Counting', *Chemical Engineering Progress*, 103(10), pp. 40–47.

Kakiuchi, M. *et al.* (2002) 'Purification, characterization, and cDNA cloning of α-N-acetylgalactosamine-specific lectin from starfish, *Asterina pectinifera'*, *Glycobiology*, 12(2), pp. 85–94.

Kennedy, J. F. *et al.* (1995) 'Lectins, versatile proteins of recognition: a review', *Carbohydrate Polymers*, 26(3), pp. 219–230.

Kim, N. S. and Lee, G. M. (2002) 'Inhibition of Sodium Butyrate-Induced Apoptosis in Recombinant Chinese Hamster Ovary Cells by Constitutively Expressing Antisense RNA of Caspase-3', *Biotechnology and Bioengineering*, 78(2), pp. 217–228. doi: 10.1002/bit.10191.

Kim, Y.-H. *et al.* (2007) 'Cloning and functional expression of the gene encoding an inhibitor against *Aspergillus flavus* α-amylase, a novel seed lectin from Lablab purpureus (Dolichos lablab)', *Plant Cell Reports*, 26(4), pp. 395–405. doi: 10.1007/s00299-006-0250-2.

Kimura, R. and Miller, W. M. (1997) 'Glycosylation of CHO-derived recombinant tPA produced under elevated pCO2', *Biotechnology Progress*, 13(3), pp. 311–317. doi: 10.1021/bp9700162.

Krasnewich, D. (2014) 'Human glycosylation disorders', *Cancer Biomarkers*, 14(1), pp. 3–16. doi: 10.3233/CBM-130374.

Kreitman, R. J. (2006) 'Immunotoxins for Targeted Cancer Therapy', *The AAPS Journal*, 8(3), pp. 532–551.

Kuystermans, D., Avesh, M. and Al-rubeai, M. (2016) 'Online flow cytometry for monitoring apoptosis in mammalian cell cultures as an application for process analytical technology', *Cytotechnology*. Springer Netherlands, 68(3), pp. 399–408. doi: 10.1007/s10616-014-9791-3.

Kuystermans, D., Mohd, A. and Al-rubeai, M. (2012) 'Automated flow cytometry for monitoring CHO cell cultures', *Methods*. Elsevier Inc., 56(3), pp. 358–365. doi: 10.1016/j.ymeth.2012.03.001.

Laemmli, U. K. (1970) 'Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4', *Nature*, 227(5259), pp. 680–685.

Lam, S. K. and Ng, T. B. (2011) 'Lectins: production and practical applications', *Applied Microbiology and Biotechnology*, 89(1), pp. 45–55. doi: 10.1007/s00253-010-2892-9.

Lannoo, N. and Damme, E. J. M. Van (2010) 'Nucleocytoplasmic plant lectins', *Biochimica et Biophysica Acta*. Elsevier B.V., 1800(2), pp. 190–201. doi: 10.1016/j.bbagen.2009.07.021.

Lee, J. *et al.* (2001) 'Chinese hamster ovary (CHO) cells may express six β4-galactosyltransferases (β4GalTs). Consequences of the loss of functional β4GalT-1, β4GalT-6, or both in CHO glycosylation mutants.', *The Journal of Biological Chemistry*, 276(17), pp. 13924–13934. doi: 10.1074/jbc.M010046200.

Lefebvre, T. *et al.* (2004) 'Modulation of O-GlcNAc glycosylation during xenopus oocyte maturation', *Journal of Cellular Biochemistry*, 93(5), pp. 999–1010. doi: 10.1002/jcb.20242.

Lepenies, B. and Seeberger, P. H. (2014) 'Simply better glycoproteins.', *Nature Biotechnology*. Nature Publishing Group, 32(5), pp. 443–5. doi: 10.1038/nbt.2893.

Ling, L.-J., Yang, Y.-Z. and Bi, Y.-R. (2010) 'Expression and characterization of two domains of Pinellia ternata agglutinin (PTA), a plant agglutinin from *Pinellia ternata* with antifungal

activity', *World Journal of Microbiology and Biotechnology*, 26(3), pp. 545–554. doi: 10.1007/s11274-009-0204-2.

Liu, B., Bian, H. and Bao, J. (2010) 'Plant lectins: Potential antineoplastic drugs from bench to clinic', *Cancer Letters*. Elsevier Ireland Ltd, 287(1), pp. 1–12. doi: 10.1016/j.canlet.2009.05.013.

Longstaff, M. *et al.* (1998) 'Production and purification of active snowdrop lectin in *Escherichia coli*', *European Journal of Biochemistry*, 252(1), pp. 59–65.

Love, D. C. and Hanover, J. A. (2005) 'The Hexosamine Signaling Pathway: Deciphering the "O-GlcNAc Code"', *Science Signaling*, 2005, pp. 1–14. doi: 10.1126/stke.3122005re13.

Luo, S., Zhangsun, D. and Tang, K. (2005) 'Functional GNA expressed in *Escherichia coli* with high efficiency and its effect on Ceratovacuna lanigera Zehntner', *Applied Microbiology and Biotechnology*, 69(2), pp. 184–191. doi: 10.1007/s00253-005-0042-6.

Ly, T. *et al.* (2017) 'Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic subphases, using PRIMMUS', *eLife*, 6, pp. 1–35. doi: 10.7554/elife.27574.

Maenuma, K. *et al.* (2008) 'Use of a library of mutated *Maackia amurensis* hemagglutinin for profiling the cell lineage and differentiation', *Proteomics*, 8(16), pp. 3274–3283. doi: 10.1002/pmic.200800037.

Marth, J. D. and Grewal, P. K. (2008) 'Mammalian glycosylation in immunity.', *Nature reviews. Immunology*, 8(11), pp. 874–87. doi: 10.1038/nri2417.

Masters, J. R. W. (2000) *Animal Cell Culture*. 3rd edn. Edited by J. R. W. Masters. New York: Oxford.

McCoy, J. P. J., Varani, J. and Goldstein, I. J. (1983) 'Enzyme-Linked Lectin Assay (ELLA): Use of Alkaline Phosphatase-Conjugated *Griffonia simplicifolia* B4 lsolectin for the Detection of a-D-galactopyranosyl End Groups', *Analytical Biochemistry*, 130(2), pp. 437–444.

McQueen, A. and Bailey, J. E. (1990) 'Effect of ammonium ion and extracellular pH on hybridoma cell metabolism and antibody production', *Biotechnology and Bioengineering*,

35(11), pp. 1067–1077. doi: 10.1002/bit.260351102.

Meinelt, E. *et al.* (2012) 'Standardizing Application Setup Across Multiple Flow Cytometers Using BD FACSDiva™ Version 6 Software', *BD Biosciences Technical Bulletin*, (March), pp. 1–16.

De Mejía, E. G. and Prisecaru, V. I. (2005) 'Lectins as Bioactive Plant Proteins: A Potential in Cancer Treatment', *Critical Reviews in Food Science and Nutrition*, 45(6), pp. 425–445. doi: 10.1080/10408390591034445.

Michiels, K., Van Damme, E. J. and Smagghe, G. (2010) 'PLANT-INSECT INTERACTIONS: WHAT CAN WE LEARN FROM PLANT LECTINS?', *Archives of Insect Biochemistry Physiology*, 73(4), pp. 193–212. doi: 10.1002/arch.20351.

Mislovičová, D. *et al.* (2009) 'Lectinomics I. Relevance of exogenous plant lectins in biomedical diagnostics', *Biologia*, 64(1), pp. 1–19. doi: 10.2478/s11756-009-0029-3.

Mitra, N. *et al.* (2006) 'N-linked oligosaccharides as outfitters for glycoprotein folding, form and function', *Trends in Biochemical Sciences*, 31(3), pp. 156–163. doi: 10.1016/j.tibs.2006.01.003.

Mody, R., Joshi, S. and Chaney, W. (1995) 'Use of Lectins as Diagnostic and Therapeutic Tools for Cancer', *Journal of Pharmacological and Toxicological Methods*, 33(1), pp. 1–10.

Moloney, D. J. *et al.* (2000) 'Mammalian Notch1 Is Modified with Two Unusual Forms of O-Linked Glycosylation Found on Epidermal Growth Factor-like Modules', *The Journal of Biological Chemistry*, 275(13), pp. 9604–9611.

Montante, S. and Brinkman, R. R. (2019) 'Flow cytometry data analysis: Recent tools and algorithms', *International Journal of Laboratory Hematology*, 41(S1), pp. 56–62. doi: 10.1111/ijlh.13016.

Muhlenhoff, M. *et al.* (1996) 'Autocatalytic polysialylation of polysialyltransferase-1', *The EMBO Journal*, 15(24), pp. 6943–6950.

Nam, J. H. *et al.* (2008) 'The Effects of Culture Conditions on the Glycosylation of Secreted

Human Placental Alkaline Phosphatase Produced in Chinese Hamster Ovary Cells', *Biotechnology and Bioengineering*, 100(6), pp. 1178–1192. doi: 10.1002/bit.21853.The.

Nathan, C. (2008) 'Epidemic Inflammation: Pondering Obesity', *Molecular Medicine*, 14(7–8), pp. 485–492. doi: 10.2119/2008-00038.nathan.

Nunez, R. (2001) 'DNA measurement and cell cycle analysis by flow cytometry.', *Current issues in molecular biology*, 3(3), pp. 67–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11488413.

Ofek, I., Hasty, D. L. and Sharon, N. (2003) 'Anti-adhesion therapy of bacterial diseases : prospects and problems', *FEMS Immunology and Medical Microbiology*, 38(3), pp. 181–191. doi: 10.1016/S0928-8244(03)00228-1.

Ohtsubo, K. and Marth, J. D. (2006) 'Glycosylation in Cellular Mechanisms of Health and Disease', *Cell*, 126(5), pp. 855–867. doi: 10.1016/j.cell.2006.08.019.

Olausson, J. *et al.* (2011) 'Production and characterization of a monomeric form and a single-site form of Aleuria aurantia lectin', *Glycobiology*, 21(1), pp. 34–44. doi: 10.1093/glycob/cwq129.

Oliveira, C. *et al.* (2009) 'cDNA Cloning and Functional Expression of the a - D -Galactose-Binding Lectin Frutalin in E*scherichia coli*', *Molecular Biotechnology*, 43, pp. 212–220. doi: 10.1007/s12033-009-9191-7.

Oliveira, C., Texeira, J. A. and Domingues, L. (2013) 'Recombinant lectins: an array of tailor-made glycan-interaction biosynthetic tools', *Critical Reviews in Biotechnology*, 33(1), pp. 66–80. doi: 10.3109/07388551.2012.670614.

Ozlu, N. *et al.* (2015) 'Quantitative comparison of a human cancer cell surface proteome between interphase and mitosis', *The EMBO Journal*, 34(2), pp. 251–265. doi: 10.15252/embj.201385162.

Park, D. *et al.* (2015) 'Characteristic Changes in Cell Surface Glycosylation Accompany Intestinal Epithelial Cell (IEC) Differentiation : High Mannose Structures Dominate the Cell

Surface Glycome of Undifferentiated Enterocytes', *Molecular & Cellular Proteomics: MCP*, 14(11), pp. 2910–2921. doi: 10.1074/mcp.M115.053983.

Patnaik, S. K. and Stanley, P. (2005) 'Mouse Large Can Modify Complex N- and Mucin O-glycans on alpha-dystroglycan to Induce Laminin Binding', *The Journal of Biological Chemistry*, 280(21), pp. 20851–20859. doi: 10.1074/jbc.M500069200.

Patsos, G. *et al.* (2009) 'O-Glycan inhibitors generate aryl-glycans, induce apoptosis and lead to growth inhibition in colorectal cancer cell lines', *Glycobiology*, 19(4), pp. 382–398. doi: 10.1093/glycob/cwn149.

Peumans, W. J. and Van Damme, E. J. M. (1995) 'Lectins as Plant Defense Proteins', *Plant Physiology*, 109(2), pp. 347–352.

Pierce, M. and Arango, J. (1986) 'Rous sarcoma virus-transformed baby hamster kidney cells express higher levels of asparagine-linked tri- and tetraantennary glycopeptides containing [GlcNAc-beta (1,6)Man-alpha (1,6)Man] and poly-N-acetyllactosamine sequences than baby hamster kidney cell', *The Journal of Biological Chemistry*, 261(23), pp. 10772–10777.

Pilobello, K. T., Slawek, D. E. and Mahal, L. K. (2007) 'A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome', *Proceedings of the National Academy of Sciences*, 104(28), pp. 11534–11539. doi: 10.1073/pnas.0704954104.

Plattner, V. E. *et al.* (2008) 'Targeted drug delivery: Binding and uptake of plant lectins using human 5637 bladder cancer cells', *European Journal of Pharmaceutics and Biopharmaceutics*, 70(2), pp. 572–576. doi: 10.1016/j.ejpb.2008.06.004.

Pozarowski, P. and Darzynkiewicz, Z. (2004) 'Analysis of the cell cycle by flow cytometry', *Methods in Molecular Biology*, 281, pp. 301–311. doi: 10.1002/bmb.41.

Propheter, D. C. and Mahal, L. K. (2011) 'Orientation of GST-tagged lectins via in situ surface modification to create an expanded lectin microarray for glycomic analysis', *Molecular BioSystems*, 7(7), pp. 2114–2117. doi: 10.1039/c1mb05047h.

Puck, T. T., Cieciura, S. J. and Robinson, A. (1958) 'Genetics of somatic mammalian cells III.

Long-term cultivation of euploid cells from human and animal subjects', *The Journal of Experimental Medicine*, 108(6), pp. 945–956.

Rahaie, M. and Kazemi, S. S. (2010) 'Lectin-based Biosensors: As Powerful Tools in Bioanalytical Applications', *Biotechnology*, 9(4), pp. 428–443.

Rahim, A. *et al.* (2018) 'High throughput automated analysis of big flow cytometry data', *Methods*, 134–135, pp. 164–176. doi: 10.1016/j.ymeth.2017.12.015.

Roberts, C. K. and Sindhu, K. K. (2009) 'Oxidative stress and metabolic syndrome', *Life Sciences*. Elsevier Inc., 84(21–22), pp. 705–712. doi: 10.1016/j.lfs.2009.02.026.

Rosenfeld, R. *et al.* (2007) 'A lectin array-based methodology for the analysis of protein glycosylation', *Journal of Biochemical and Biophysical Methods*, 70(3), pp. 415–426. doi: 10.1016/j.jbbm.2006.09.008.

Rosner, B. (2000) *Fundamentals of Biostatistics*. 5th edn. Edited by L. Campobasso. Duxburry, pp. 273 - 312.

Röttger, S. *et al.* (1998) 'Localization of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus', *Journal of Cell Science*, 111(Pt 1), pp. 45–60.

Ruddock, L. W. and Molinari, M. (2006) 'N-glycan processing in ER quality control', *Journal of Cell Science*, 119(Pt21), pp. 4373–4380. doi: 10.1242/jcs.03225.

Sasaki, H. *et al.* (1987) 'Carbohydrate Structure of Erythropoietin Expressed in Chinese Hamster Ovary Cells by a Human Erythropoietin cDNA', *The Journal of Biological Chemistry*, 262(25), pp. 12059–12076.

Schachter, H. and Brockhausen, I. (1992) 'The biosynthesis of serine (threonine)-N-acetylgalactosamine-linked carbohydrate moieties.', in Allen, H. . and Kisailus, E. . (eds) *Glycoconjugates: composition, structure and function*. New York: Marcel Dekker, pp. 262–332.

Shapiro, H. M. and Telford, W. G. (2009) 'Lasers for Flow Cytometry', *Curr. Protoc. Cytom.*, pp. 1–17. doi: 10.1002/0471142956.cy0109s49.

Sharon, N. and Lis, H. (1989) 'Lectins as Cell Recognition Molecules', *Science*, 246(4927), pp. 227–234.

Sharon, N. and Lis, H. (1993) 'Carbohydrates in Cell Recognition', *Scientific American*, 268(January), pp. 82–89.

Shi, H. H. and Goudar, C. T. (2014) 'Recent advances in the understanding of biological implications and modulation methodologies of monoclonal antibody N-linked high mannose glycans', *Biotechnology and Bioengineering*, 111(10), pp. 1907–1919. doi: 10.1002/bit.25318.

Sitton, G. and Srienc, F. (2008) 'Mammalian cell culture scale-up and fed-batch control using automated flow cytometry', *Journal of Biotechnology*, 135(2), pp. 174–180. doi: 10.1016/j.jbiotec.2008.03.019.

Slawson, C. *et al.* (2005) 'Perturbations in O-linked β-N-acetylglucosamine protein modification cause severe defects in mitotic progression and cytokinesis', *Journal of Biological Chemistry*, 280(38), pp. 32944–32956. doi: 10.1074/jbc.M503396200.

Slivac, I. *et al.* (2010) 'Influence of different ammonium, lactate and glutamine concentrations on CCO cell growth', *Cytotechnology*, 62(6), pp. 585–594. doi: 10.1007/s10616-010-9312-y.

Sørensen, H. P. and Mortensen, K. K. (2005) 'Advanced genetic strategies for recombinant protein expression in *Escherichia coli*', *Journal of Biotechnology*, 115(2), pp. 113–128. doi: 10.1016/j.jbiotec.2004.08.004.

Sou, S. N. *et al.* (2015) 'How does mild hypothermia affect monoclonal antibody glycosylation?', *Biotechnology and Bioengineering*, 112(6), pp. 1165–1176. doi: 10.1002/bit.25524.

Stancombe, P. R. *et al.* (2003) 'Isolation of the gene and large-scale expression and purification of recombinant *Erythrina cristagalli* lectin', *Protein Expression and Purification*, 30(2), pp. 283–292. doi: 10.1016/S1046-5928(03)00125-6.

Stanley, P., Sudo, T. and Carver, J. P. (1980) 'Differential involvement of cell surface sialic acid residues in wheat germ agglutinin binding to parental and wheat germ agglutinin-resistant

Chinese hamster ovary cells', *The Journal of Cell Biology*, 85(1), pp. 60–69.

Stanley, P. and Sundaram, S. (2014) 'Rapid Assays for Lectin Toxicity and Binding Changes that Reflect Altered Glycosylation in Mammalian Cells', 6, pp. 117–133. doi: 10.1002/9780470559277.ch130206.

Streicher, H. and Sharon, N. (2003) 'Recombinant Plant Lectins and Their Mutants', *Methods in Enzymology*, 363, pp. 47–77.

Subtelny, S. and Wessells, N. K. (eds) (1980) *The Cell Surface: Mediator of Developmental Processes*. London: Academic Press.

Sung, Y. H. *et al.* (2004) 'Effect of sodium butyrate on the production, heterogeneity and biological activity of human thrombopoietin by recombinant Chinese hamster ovary cells', *Journal of Biotechnology*, 112(3), pp. 323–335. doi: 10.1016/j.jbiotec.2004.05.003.

Tate, S. and Ko Ferrigno, P. (2006) 'Cell Cycle: Synchronization at Various Stages', *Encyclopedia of Life Sciences*, 1, pp. 1–5. doi: 10.1038/npg.els.0002570.

Tateno, H. *et al.* (2007) 'A novel strategy for mammalian cell surface glycome profiling using lectin microarray', *Glycobiology*, 17(10), pp. 1138–1146. doi: 10.1093/glycob/cwm084.

Taylor, M. and Drickamer, K. (2006) *Introduction to Glycobiology*. 2nd edn. New York: Oxford University Press Inc.

Teixeira, A. P. *et al.* (2009) 'Advances in on-line monitoring and control of mammalian cell cultures: Supporting the PAT initiative', *Biotechnology Advances*. Elsevier Inc., 27(6), pp. 726–732. doi: 10.1016/j.biotechadv.2009.05.003.

Thompson, R. *et al.* (2011) 'Optimization of the enzyme-linked lectin assay for enhanced glycoprotein and glycoconjugate analysis', *Analytical Biochemistry*. Elsevier Inc., 413(2), pp. 114–122. doi: 10.1016/j.ab.2011.02.013.

Thorens, B. and Vassalli, P. (1986) 'Chloroquine and ammonium chloride prevent terminal glycosylation of immunoglobulins in plasma cells without affecting secretion', *Nature*, 321(6070), pp. 618–620. doi: 10.1038/321618a0.

Tielker, D. *et al.* (2006) 'Lectin-based affinity tag for one-step protein purification', *BioTechniques*, 41(3), pp. 327–332. doi: 10.2144/000112236.

Trachootham, D., Alexandre, J. and Huang, P. (2009) 'Targeting cancer cells by ROS-mediated mechanisms: A radical therapeutic approach?', *Nature Reviews Drug Discovery*. Nature Publishing Group, 8(7), pp. 579–591. doi: 10.1038/nrd2803.

Tremblay, R. *et al.* (2011) 'High-yield expression of recombinant soybean agglutinin in plants using transient and stable systems', *Transgenic Research*, 20(2), pp. 345–356. doi: 10.1007/s11248-010-9419-0.

Underhill, M. F. and Smales, C. M. (2007) 'The cold-shock response in mammalian cells: Investigating the HeLa cell cold-shock proteome', *Cytotechnology*, 53(1–3), pp. 47–53. doi: 10.1007/s10616-007-9048-5.

Upadhyay, S. K. *et al.* (2010) 'SUMO fusion facilitates expression and purification of garlic leaf lectin but modifies some of its properties', *Journal of Biotechnology*. Elsevier B.V., 146(1–2), pp. 1–8. doi: 10.1016/j.jbiotec.2010.01.013.

Varki, A. and Sharon, N. (2009) 'Historical Background and Overview', in Varki, A. et al. (eds) *Essentials of glycobiology*. 2nd edn. New York: Cold Spring Harbor Laboratory Press, pp. 1 -22.

Veal, E. A., Day, A. M. and Morgan, B. A. (2007) 'Hydrogen peroxide sensing and signaling', *Molecular Cell*, 26, pp. 1–14. doi: 10.1007/978-81-322-2035-0_8.

Veiseh, M. *et al.* (2014) 'Cellular heterogeneity profiling by hyaluronan probes reveals an invasive but slow-growing breast tumor subset', *Proceedings of the National Academy of Sciences of the United States of America*, 111(17), pp. 1731–1739. doi: 10.1073/pnas.1402383111.

Wada, R., Matsui, M. and Kawasaki, N. (2019) 'Influence of N-glycosylation on effector functions and thermal stability of glycoengineered IgG1 monoclonal antibody with homogeneous glycoforms', *mAbs*. Taylor & Francis, 11(2), pp. 350–372. doi: 10.1080/19420862.2018.1551044.

Wade, J. L. G. (1999) *Organic Chemistry*. 4th edn. New Jersey: Prentice-Hall Inc.

Wallace, D. C. (2005) 'A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine.', *Annual review of genetics*, 39, pp. 359–407. doi: 10.1146/annurev.genet.39.110304.095751.

Walsh, C. T., Garneau-tsodikova, S. and Gatto, G. J. (2005) 'Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications', *Angewandte Chemie (International Edition in English)*, 44(45), pp. 7342–7372. doi: 10.1002/anie.200501023.

Warnock, D. E. *et al.* (1993) 'Determination of Plasma Membrane Lipid Mass and Composition in Cultured Chinese Hamster Ovary Cells Using High Gradient Magnetic Affinity Chromatography', *The Journal of Biological Chemistry*, 268(14), pp. 10145–10153.

Wellen, K. E. and Thompson, C. B. (2010) 'Cellular Metabolic Stress: Considering How Cells Respond to Nutrient Excess', *Molecular Cell*. Elsevier Inc., 40(2), pp. 323–332. doi: 10.1016/j.molcel.2010.10.004.

Werner, R. G., Kopp, K. and Schlueter, M. (2007) 'Glycosylation of therapeutic proteins in different production systems', *Acta Paediatrica, International Journal of Paediatrics*, 96(SUPPL. 455), pp. 17–22. doi: 10.1111/j.1651-2227.2007.00199.x.

Wilkerson, M. J. (2012) 'Principles and Applications of Flow Cytometry and Cell Sorting in Companion Animal Medicine', *Veterinary Clinics of North America: Small Animal Practice*. Elsevier Inc., 42(1), pp. 53–71. doi: 10.1016/j.cvsm.2011.09.012.

Woo, S. A. *et al.* (2008) 'Effect of culture temperature on erythropoietin production and glycosylation in a perfusion culture of recombinant CHO cells', *Biotechnology and Bioengineering*, 101(6), pp. 1234–1244. doi: 10.1002/bit.22006.

Yabe, R. *et al.* (2007) 'Tailoring a Novel Sialic Acid-Binding Lectin from a Ricin-B Chain-like Galactose-Binding Protein by Natural Evolution-Mimicry', *Journal of Biochemistry*, 141(3), pp. 389–399. doi: 10.1093/jb/mvm043.

Yamashita, K. *et al.* (1985) 'Enzymatic Basis for the Structural Changes of Asparagine-linked

Sugar Chains of Membrane Glycoproteins of Baby Hamster Kidney', *The Journal of Biological Chemistry*, 260(April), pp. 3963–3969.

Yang, N. *et al.* (2005) 'Molecular Character of the Recombinant Antitumor Lectin from the Edible Mushroom *Agrocybe aegerita*', *Journal of Biochemistry*, 138(2), pp. 145–150. doi: 10.1093/jb/mvi109.

Yerganian, G. (1972) 'Pathology of the Syrian Hamster', *Progress in Experimental Tumor Research*, 16, pp. 2–41.

Yerganian, G. (1985) 'The biology and genetics of Chinese hamster', *Molecular Cell Genetics*, edited by Michael Gottesman, New York: John Wiley & Sons, pp. 3–36.

Yim, M., Ono, T. and Irimura, T. (2001) 'Mutated plant lectin library useful to identify different cells', *Proceedings of the National Academy of Sciences of the United States of America*, 98(5), pp. 2222–2225.

Yin, J. *et al.* (2007) 'Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes', *Journal of Biotechnology*, 127(3), pp. 335–347. doi: 10.1016/j.jbiotec.2006.07.012.

Zachara, N. E. and Hart, G. W. (2006) 'Cell signaling, the essential role of O-GlcNAc!', *Biochimica et Biophysica Acta 1761*, 1761(5–6), pp. 599–617. doi: 10.1016/j.bbalip.2006.04.007.

Zanghi, J. A. *et al.* (1999) 'Bicarbonate concentration and osmolality are key determinants in the inhibition of CHO cell polysialylation under elevated pCO2 or pH', *Biotechnology & Bioengineering*, 65(2), pp. 182–191. doi: 10.1002/(sici)1097-0290(19991020)65:2<182::aid-bit8>3.3.co;2-4.

Zhang, K. and Kaufman, R. J. (2006) 'The unfolded protein response: A stress signaling pathway critical for health and disease', *Neurology*, 66(2 Suppl 1), pp. S102–S109. doi: 10.1212/01.wnl.0000192306.98198.ec.

Zhao, L. *et al.* (2015) 'Advances in process monitoring tools for cell culture bioprocesses',

Engineering in Life Sciences, 15(5), pp. 459–468. doi: 10.1002/elsc.201500006.

Zhao, R., Natarajan, A. and Srienc, F. (1999) 'A Flow Injection Flow Cytometry System for On-Line Monitoring of Bioreactors', *Biotechnology and Bioengineering*, 62(5), pp. 609–617.

Zheng, K. *et al.* (2014) 'Influence of glycosylation pattern on the molecular properties of monoclonal antibodies', *mAbs*, 6(3), pp. 649–658. doi: 10.4161/mabs.28588.

Zheng, K., Bantog, C. and Bayer, R. (2011) 'The impact of glycosylation on monoclonal antibody conformation and stability', *mAbs*, 3(6). doi: 10.4161/mabs.3.6.17922.

8 Appendix

8.1 Creation of functions

This section comprises of the creation of functions specifically designed to treat the raw data extracted from the flow cytometer. These functions filter, organize and perform statistical analysis of the data.

```{r}
source("http://bioconductor.org/biocLite.R")
biocLite("flowWorkspace")
install.packages("stringi")
library(flowCore)
library(flowWorkspace)
library(ggplot2)
library(ggcyto)
library(dplyr)
library(stringr)
library(openxlsx)
```

Read in the files on R from a file

Files from the replicates (triplicate experiment) are organized into 3 lists containing the application settings control and the full stained samples.

```{r}
flow_gating <- function(x, x_WGA) {
setwd(x)
fclist1 <- c("Application Settings_Unstained.fcs", "Application
Settings_7AAD.fcs","Application Settings_D5.fcs", "Application Settings_AAL.fcs",
"Application Settings_AAL-2.fcs", "Application Settings_MAL II.fcs", "Application
Settings_PNA.fcs", "Application Settings_WGA.fcs", "Application Settings_LECA.fcs",
"Application Settings_LECB.fcs", "Samples_AAL I.fcs", "Samples_AAL-2 I.fcs", "Samples_MAL
```

```
II I.fcs" , "Samples_PNA I.fcs" , "Samples_WGA I.fcs", "Samples_LECA I.fcs" , "Samples_LECB
I.fcs")

fclist2 <- c("Application Settings_Unstained.fcs", "Application
Settings_7AAD.fcs","Application Settings_D5.fcs", "Application Settings_AAL.fcs",
"Application Settings_AAL-2.fcs", "Application Settings_MAL II.fcs", "Application
Settings_PNA.fcs", "Application Settings_WGA.fcs", "Application Settings_LECA.fcs",
"Application Settings_LECB.fcs", "Samples_AAL II.fcs", "Samples_AAL-2 II.fcs",
"Samples_MAL II II.fcs" , "Samples_PNA II.fcs" , "Samples_WGA II.fcs", "Samples_LECA II.fcs"
, "Samples_LECB II.fcs")

 fclist3 <- c("Application Settings_Unstained.fcs", "Application
Settings_7AAD.fcs","Application Settings_D5.fcs", "Application Settings_AAL.fcs",
"Application Settings_AAL-2.fcs", "Application Settings_MAL II.fcs", "Application
Settings_PNA.fcs", "Application Settings_WGA.fcs", "Application Settings_LECA.fcs",
"Application Settings_LECB.fcs", "Samples_AAL III.fcs", "Samples_AAL-2 III.fcs",
"Samples_MAL II III.fcs" , "Samples_PNA III.fcs" , "Samples_WGA III.fcs", "Samples_LECA
III.fcs" , "Samples_LECB III.fcs")

#Creation of flowsets containing the unstransformed files of the lists.

fs1 <- read.flowSet(fclist1, transformation = FALSE)

fs2 <- read.flowSet(fclist2, transformation = FALSE)

fs3 <- read.flowSet(fclist3, transformation = FALSE)

#compensation matrix calculation

frames <- lapply(dir(x_WGA, full.names=TRUE), read.FCS)

names(frames) <- c("Unstained", "7AAD-A", "D5-A", "LECTIN-A")

frames <- as(frames, "flowSet")

comp <- spillover(frames, unstained = "Unstained" , patt = "-A", fsc = "FSC-A", ssc = "SSC-A",
stain_match = "ordered")

#compensation loop

for (i in 11:17) {

fs1[[i]] <- compensate(fs1[[i]], comp)

fs2[[i]] <- compensate(fs2[[i]], comp)

fs3[[i]] <- compensate(fs3[[i]], comp)

}

#Perform transformation of the parameters of PE-Texas Red.

lgcl <- estimateLogicle(fs1[[1]], channels = c("PE-Texas Red-A" ,"PE-Texas Red-H"))

after1 <- transform(fs1, lgcl)
```

```
lgcl <- estimateLogicle(fs2[[1]], channels = c("PE-Texas Red-A" ,"PE-Texas Red-H"))
after2 <- transform(fs2, lgcl)
lgcl <- estimateLogicle(fs3[[1]], channels = c("PE-Texas Red-A" ,"PE-Texas Red-H"))
after3 <- transform(fs3, lgcl)
#Creation of data hierarchy
#The transformed data is used here
gs1 <- GatingSet(after1)
gs2 <- GatingSet(after2)
gs3 <- GatingSet(after3)
#Creation of the Non debris gates for gs1, gs2 and gs3
#1. Nondebris (setting up the gate parameters)
rg1 <- rectangleGate("FSC-A"=c(50000,Inf), filterId = "NonDebris")
#Adding rg1 to gs1 and defining the rg1 parent
add(gs1, rg1, parent = "root")
#Updating gs1
recompute(gs1)
#Adding rg1 to gs2 and defining the rg1 parent
add(gs2, rg1, parent = "root")
#Updating gs2
recompute(gs2)
#Adding rg1 to gs3 and defining the rg1 parent
add(gs3, rg1, parent = "root")
#Updating gs3
recompute(gs3)
#Aggregates exclusion
#Setting up the first gate to exclude aggregates
#Visualizing FSC-A vs FSC-H plot(Pacific Blue sample - application settings files) to select
coordinates of the gate
#autoplot(after1[[4]], "FSC-A", "FSC-H")
```

```
#Setting the gate
pg2 <- polygonGate( filterId = "singletsa", cbind("FSC-A" =
c(0,50000,150000,300000,300000,0), "FSC-H" = c(0,17000,48000,62000,120000,105000)))
add(gs1,pg2, parent = "NonDebris", name = "singletsa")
recompute(gs1)
add(gs2,pg2, parent = "NonDebris", name = "singletsa")
recompute(gs2)
add(gs3,pg2, parent = "NonDebris", name = "singletsa")
recompute(gs3)
#Setting up the second gate to exclude aggregates
#This is a way to get the data to visualize the filtered events only
gs1singletsA <- getData(gs1,"/NonDebris/singletsa")
#autoplot(gs1singletsA[[4]],"FSC-A", "FSC-W")
#Setting up the gate
pg3 <- polygonGate( filterId = "singletsb", cbind("FSC-A" = c(50000,50000,275000,275000),
"FSC-W" = c(75000,112500,212500,75000)))
add(gs1,pg3, parent = "singletsa", name = "singletsb")
recompute(gs1)
add(gs2,pg3, parent = "singletsa", name = "singletsb")
recompute(gs2)
add(gs3,pg3, parent = "singletsa", name = "singletsb")
recompute(gs3)
#Seeting up the gates for Live, dead and apoptotic cells
#Visualizing the filtered data
gs1singletsonly <- getData(gs1, "/NonDebris/singletsa/singletsb")
#Setting the live gate
rg2 <- rectangleGate("PE-Texas Red-A"=c(0,2.5), "PE-Texas Red-H"=c(0,2.5), filterId = "Alive")
add(gs1, rg2, parent = "singletsb")
recompute(gs1)
add(gs2, rg2, parent = "singletsb")
```

```
recompute(gs2)
add(gs3, rg2, parent = "singletsb")
recompute(gs3)
#Setting the gate for dead cells
rg3<- rectangleGate("PE-Texas Red-A"=c(3.25,5), "PE-Texas Red-H"=c(3,5), filterId = "Dead")
add(gs1, rg3, parent = "singletsb")
recompute(gs1)
add(gs2, rg3, parent = "singletsb")
recompute(gs2)
add(gs3, rg3, parent = "singletsb")
recompute(gs3)
#Setting the gate for dead cells including apoptotic cells
rg3_deadapo <- rectangleGate("PE-Texas Red-A"=c(2.5,5), "PE-Texas Red-H"=c(2.0,5), filterId
= "DeadApo")
add(gs1, rg3_deadapo, parent = "singletsb")
recompute(gs1)
add(gs2, rg3_deadapo, parent = "singletsb")
recompute(gs2)
add(gs3, rg3_deadapo, parent = "singletsb")
recompute(gs3)
#Setting up the gate of apoptotic cells
rg4 <- rectangleGate("PE-Texas Red-A"=c(2.5,3.25), "PE-Texas Red-H"=c(2,3), filterId =
"Apoptotic")
add(gs1, rg4, parent = "singletsb")
recompute(gs1)
add(gs2, rg4, parent = "singletsb")
recompute(gs2)
add(gs3, rg4, parent = "singletsb")
recompute(gs3)
#Setting up the 1XDNA, IntDNA and 2XDNA cells
```

```
#The coordinates for the gates must be calculated based on D5 histogram (APC channel)
The coordinates are calculated for each APC sample.

#We must mathematically determine the APC-Cy7-A value for the highest peak on the
histogram. This value tells where to gate for the 1XDNA cells. Twice as much of this value
we can then gate the 2xDNA cells. The median value between 1X and 2XDNA cells is where
the intDNA cells are found.

#Gating DNA cell cycle of gs1

#Gating DNA cell cycle cells from the Alive gate

gs1alive <- getData(gs1, "/NonDebris/singletsa/singletsb/Alive")

for (i in 1:17){

test <- data.frame(exprs(gs1alive[[i]]))

n <- length((subset(density(test$APC.Cy7.A)$x, density(test$APC.Cy7.A)$x < 200000)))

ymax <- which.max(density(test$APC.Cy7.A)$y[c(1:n)])

a1<-density(test$APC.Cy7.A)$x[ymax]

a2<-a1*2

a15<-a1*1.5

x<-a1/10

rg5 <- rectangleGate("APC-Cy7-A"=c((a1-x),(a1+x)), filterId = "1xcells")

rg6 <- rectangleGate("APC-Cy7-A"=c((a15-x),(a15+x)), filterId = "int")

rg7 <- rectangleGate("APC-Cy7-A"=c((a2-x),(a2+x)), filterId = "2xcells")

add(gs1[[i]],rg5, parent ="Alive")

add(gs1[[i]],rg6, parent ="Alive")

add(gs1[[i]],rg7, parent ="Alive")

recompute(gs1[[i]])

}

#Gating DNA cell cycle of gs2

#Gating DNA cell cycle cells from the Alive gate

gs2alive <- getData(gs2, "/NonDebris/singletsa/singletsb/Alive")

for (i in 1:17){

test<- data.frame(exprs(gs2alive[[i]]))

n <- length((subset(density(test$APC.Cy7.A)$x, density(test$APC.Cy7.A)$x < 200000)))
```

```
ymax <- which.max(density(test$APC.Cy7.A)$y[c(1:n)])
a1<-density(test$APC.Cy7.A)$x[ymax]
a2<-a1*2
a15<-a1*1.5
x<-a1/10
rg5 <- rectangleGate("APC-Cy7-A"=c((a1-x),(a1+x)), filterId = "1xcells")
rg6 <- rectangleGate("APC-Cy7-A"=c((a15-x),(a15+x)), filterId = "int")
rg7 <- rectangleGate("APC-Cy7-A"=c((a2-x),(a2+x)), filterId = "2xcells")
add(gs2[[i]],rg5, parent ="Alive")
add(gs2[[i]],rg6, parent ="Alive")
add(gs2[[i]],rg7, parent ="Alive")
recompute(gs2[[i]])
}
#Gating DNA cell cycle of gs3
#Gating DNA cell cycle cells from the Alive gate
gs3alive <- getData(gs3, "/NonDebris/singletsa/singletsb/Alive")
for (i in 1:17){
test<- data.frame(exprs(gs3alive[[i]]))
n <- length((subset(density(test$APC.Cy7.A)$x, density(test$APC.Cy7.A)$x < 200000)))
ymax <- which.max(density(test$APC.Cy7.A)$y[c(1:n)])
a1<-density(test$APC.Cy7.A)$x[ymax]
a2<-a1*2
a15<-a1*1.5
x<-a1/10
rg5 <- rectangleGate("APC-Cy7-A"=c((a1-x),(a1+x)), filterId = "1xcells")
rg6 <- rectangleGate("APC-Cy7-A"=c((a15-x),(a15+x)), filterId = "int")
rg7 <- rectangleGate("APC-Cy7-A"=c((a2-x),(a2+x)), filterId = "2xcells")
add(gs3[[i]],rg5, parent ="Alive")
add(gs3[[i]],rg6, parent ="Alive")
```

```
add(gs3[[i]],rg7, parent ="Alive")
recompute(gs3[[i]])
}
#autoplot(gs1, c("1xcells", "2xcells", "int"))
flow_gating_list <- list(gs1, gs2, gs3)
flow_gating_list
}
#End of flow_gating()
# Creating a function to equalize sample sizes
equalize_samplesize <- function (x, y, z) {
  n_rows_vec <- c(nrow(x), nrow(y), nrow(z))
smallest_sample_size <- min(n_rows_vec)
x <- x[sample(nrow(x), smallest_sample_size), ]
y <- y[sample(nrow(y), smallest_sample_size), ]
z <- z[sample(nrow(z), smallest_sample_size), ]
list(x, y, z)
}
 #End of equalize_samplesize()
#Function to calculate the means of a column (all parameters) of each repeated
experiment
col_means <- function(list, names_vec) {
    output <- vector(mode = "list")
  for (j in 1:length(list)) {
    output[[j]] <- j
     for (i in 1:17) {
            output[[j]][i] <- mean(list[[j]][,i], na.rm = TRUE
  }
      output[[j]] <- rbind(c(1:17), output[[j]])
      colnames(output[[j]]) <- names_vec
```

```
237         rownames(output[[j]]) <- c("Column Number", "Mean")
238       }
239       output
240     }
241     #Function to calculate the standard deviation (all parameters) of each repeated
242     experiment
243     col_sd <- function(list, names_vec) {
244       output <- vector(mode = "list")
245       for (j in 1:length(list)) {
246         output[[j]] <- j
247         for (i in 1:17) {
248                 output[[j]][i] <- sd(list[[j]][,i], na.rm = TRUE)
249       }
250         output[[j]] <- rbind(c(1:17), output[[j]])
251         colnames(output[[j]]) <- names_vec
252         rownames(output[[j]]) <- c("Column Number", "Standard Deviation")
253       }
254       output
255     }
256     #Function to calculate the Coeffient of Variation (all parameters) of each repeated
257     experiment
258     col_cv <- function(list, names_vec) {
259       output <- vector(mode = "list")
260       for (j in 1:length(list)) {
261         output[[j]] <- j
262         for (i in 1:17) {
263           output[[j]][i] <- 100 * (sd(list[[j]][,i], na.rm = TRUE) / mean(list[[j]][,i], na.rm = TRUE))
264         }
265         output[[j]] <- rbind(c(1:17), output[[j]])
266         colnames(output[[j]]) <- names_vec
```

```
      rownames(output[[j]]) <- c("Column Number", "Coefficient of Variation")
  }
  output
}
```

#Calculation of the mean of the means, means of the standard deviations and the means of the coefficient of variation

##Creating the function to calculate the mean of the means of biological replicates

```
col_means2 <- function(list, names_vec) {
   output <- numeric(17)
     for (i in 1:17) {
     output[[i]] <- mean(c(list[[1]][2,i], list[[2]][2,i], list[[3]][2,i]), na.rm = TRUE)
    }
     output <- rbind(c(1:17), output)
     colnames(output) <- names_vec
     rownames(output) <- c("Column Number", "Mean of the Samples Means")
     output <- output[-c(1),]
output
}
```

##Creating the function to calculate the mean of the standard deviations of biological replicates

```
col_sd2 <- function(list, names_vec) {
   output <- numeric(17)
     for (i in 1:17) {
     output[[i]] <- mean(c(list[[1]][2,i], list[[2]][2,i], list[[3]][2,i]), na.rm = TRUE)
    }
     output <- rbind(c(1:17), output)
     colnames(output) <- names_vec
     rownames(output) <- c("Column Number", "Mean of the Sd's")
     output <- output[-c(1),]
output
```

```
}
##Creating the function to calculate the mean of the coefficient of variation of biological replicates
col_cv2 <- function(list, names_vec) {
   output <- numeric(17)
     for (i in 1:17) {
     output[[i]] <- mean(c(list[[1]][2,i], list[[2]][2,i], list[[3]][2,i]), na.rm = TRUE)
    }
     output <- rbind(c(1:17), output)
     colnames(output) <- names_vec
     rownames(output) <- c("Column Number", "Mean of CV's")
     output <- output[-c(1),]
output
}
#Function to create a data frame containing the data from an experiment (Triplicates combined)
table_summary <- function(x, y, z, sample_type) {
#Extracting flow data in variables
gs1Alive1x  <- getData(gs1,"/NonDebris/singletsa/singletsb/Alive/1xcells")
gs1AliveInt  <- getData(gs1,"/NonDebris/singletsa/singletsb/Alive/int")
gs1Alive2x  <- getData(gs1,"/NonDebris/singletsa/singletsb/Alive/2xcells")
gs1Dead  <- getData(gs1,"/NonDebris/singletsa/singletsb/Dead")
gs1Apo <- getData(gs1, "/NonDebris/singletsa/singletsb/Apoptotic")
gs1DeadApo <- getData(gs1, "/NonDebris/singletsa/singletsb/DeadApo")
gs1Alive <- getData(gs1,"/NonDebris/singletsa/singletsb/Alive")
gs2Alive1x  <- getData(gs2,"/NonDebris/singletsa/singletsb/Alive/1xcells")
gs2AliveInt  <- getData(gs2,"/NonDebris/singletsa/singletsb/Alive/int")
gs2Alive2x  <- getData(gs2,"/NonDebris/singletsa/singletsb/Alive/2xcells")
gs2Dead  <- getData(gs2,"/NonDebris/singletsa/singletsb/Dead")
gs2Apo<- getData(gs2, "/NonDebris/singletsa/singletsb/Apoptotic")
```

```
gs2DeadApo <- getData(gs2, "/NonDebris/singletsa/singletsb/DeadApo")

gs2Alive <- getData(gs2,"/NonDebris/singletsa/singletsb/Alive")

gs3Alive1x  <- getData(gs3,"/NonDebris/singletsa/singletsb/Alive/1xcells")

gs3AliveInt  <- getData(gs3,"/NonDebris/singletsa/singletsb/Alive/int")

gs3Alive2x  <- getData(gs3,"/NonDebris/singletsa/singletsb/Alive/2xcells")

gs3Dead  <- getData(gs3,"/NonDebris/singletsa/singletsb/Dead")

gs3Apo<- getData(gs3, "/NonDebris/singletsa/singletsb/Apoptotic")

gs3DeadApo <- getData(gs3, "/NonDebris/singletsa/singletsb/DeadApo")

gs3Alive <- getData(gs3,"/NonDebris/singletsa/singletsb/Alive")

#Lists for the contruction of the data frame containing all the information processed across
the 6 filters(subpopulations) of each file(lectins)

lectin_loop_list <- vector(mode = "list")

filter_loop_list <- vector(mode = "list")

rbinding_filter_loop_list <- vector(mode = "list")

#Vector to label the flow cytometry channels, lectins and subpopulations identified

names_vector <- c("FSC-A", "FSC-H", "FSC-W", "SSC-A", "SSC-H", "SSC-W", "7AAD-A", "7AAD-
H", "7AAD-W", "DRAQ5-A", "DRAQ5-H", "DRAQ5-W", "LECTIN-A", "LECTIN-H", "LECTIN-W",
"Area_ratio", "Height_ratio")

lectin_names_vec <- as.data.frame(c("AAL", "AAL-2", "MAL II", "PNA", "WGA", "LEC A", "LEC
B"), stringsAsFactors = FALSE)

filter_names_vec <- c("G2/M", "S", "Go/G1", "Dead", "Apoptotic", "Dead + Apoptotic")

#Loop to iterate through the lectin files (from 11th to 17th file)

for (lec in 11:17) {

#Extracting the data in a data frame format

gs1_Alive_1x_df <- data.frame(exprs(gs1Alive1x[[lec]]))

gs1_Alive_1x_df <- mutate(gs1_Alive_1x_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio
= Pacific.Blue.H/FSC.H)

gs1_Alive_1x_df <- gs1_Alive_1x_df[, -16]

gs2_Alive_1x_df <- data.frame(exprs(gs2Alive1x[[lec]]))

gs2_Alive_1x_df <- mutate(gs2_Alive_1x_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio
= Pacific.Blue.H/FSC.H)
```

```
gs2_Alive_1x_df <- gs2_Alive_1x_df[, -16]

gs3_Alive_1x_df <- data.frame(exprs(gs3Alive1x[[lec]]))

gs3_Alive_1x_df <- mutate(gs3_Alive_1x_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio
= Pacific.Blue.H/FSC.H)

gs3_Alive_1x_df <- gs3_Alive_1x_df[, -16]

gs1_Alive_int_df <- data.frame(exprs(gs1AliveInt[[lec]]))

gs1_Alive_int_df <- mutate(gs1_Alive_int_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio
= Pacific.Blue.H/FSC.H)

gs1_Alive_int_df <- gs1_Alive_int_df[, -16]

gs2_Alive_int_df <- data.frame(exprs(gs2AliveInt[[lec]]))

gs2_Alive_int_df <- mutate(gs2_Alive_int_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio
= Pacific.Blue.H/FSC.H)

gs2_Alive_int_df <- gs2_Alive_int_df[, -16]

gs3_Alive_int_df <- data.frame(exprs(gs3AliveInt[[lec]]))

gs3_Alive_int_df <- mutate(gs3_Alive_int_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio
= Pacific.Blue.H/FSC.H)

gs3_Alive_int_df <- gs3_Alive_int_df[, -16]

gs1_Alive_2x_df <- data.frame(exprs(gs1Alive2x[[lec]]))

gs1_Alive_2x_df <- mutate(gs1_Alive_2x_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio
= Pacific.Blue.H/FSC.H)

gs1_Alive_2x_df <- gs1_Alive_2x_df[, -16]

gs2_Alive_2x_df <- data.frame(exprs(gs2Alive2x[[lec]]))

gs2_Alive_2x_df <- mutate(gs2_Alive_2x_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio
= Pacific.Blue.H/FSC.H)

gs2_Alive_2x_df <- gs2_Alive_2x_df[, -16]

gs3_Alive_2x_df <- data.frame(exprs(gs3Alive2x[[lec]]))

gs3_Alive_2x_df <- mutate(gs3_Alive_2x_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio
= Pacific.Blue.H/FSC.H)

gs3_Alive_2x_df <- gs3_Alive_2x_df[, -16]

gs1_Dead_df <- data.frame(exprs(gs1Dead[[lec]]))

gs1_Dead_df <- mutate(gs1_Dead_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio =
Pacific.Blue.H/FSC.H)
```

```
gs1_Dead_df <- gs1_Dead_df[, -16]

gs2_Dead_df <- data.frame(exprs(gs2Dead[[lec]]))

gs2_Dead_df <- mutate(gs2_Dead_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio =
Pacific.Blue.H/FSC.H)

gs2_Dead_df <- gs2_Dead_df[, -16]

gs3_Dead_df <- data.frame(exprs(gs3Dead[[lec]]))

gs3_Dead_df <- mutate(gs3_Dead_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio =
Pacific.Blue.H/FSC.H)

gs3_Dead_df <- gs3_Dead_df[, -16]

gs1_Apo_df <- data.frame(exprs(gs1Apo[[lec]]))

gs1_Apo_df <- mutate(gs1_Apo_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio =
Pacific.Blue.H/FSC.H)

gs1_Apo_df <- gs1_Apo_df[, -16]

gs2_Apo_df <- data.frame(exprs(gs2Apo[[lec]]))

gs2_Apo_df <- mutate(gs2_Apo_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio =
Pacific.Blue.H/FSC.H)

gs2_Apo_df <- gs2_Apo_df[, -16]

gs3_Apo_df <- data.frame(exprs(gs3Apo[[lec]]))

gs3_Apo_df <- mutate(gs3_Apo_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio =
Pacific.Blue.H/FSC.H)

gs3_Apo_df <- gs3_Apo_df[, -16]

gs1_DeadApo_df <- data.frame(exprs(gs1DeadApo[[lec]]))

gs1_DeadApo_df <- mutate(gs1_DeadApo_df, Area_ratio = Pacific.Blue.A/FSC.A,
Height_ratio = Pacific.Blue.H/FSC.H)

gs1_DeadApo_df <- gs1_DeadApo_df[, -16]

gs2_DeadApo_df <- data.frame(exprs(gs2DeadApo[[lec]]))

gs2_DeadApo_df <- mutate(gs2_DeadApo_df, Area_ratio = Pacific.Blue.A/FSC.A,
Height_ratio = Pacific.Blue.H/FSC.H)

gs2_DeadApo_df <- gs2_DeadApo_df[, -16]

gs3_DeadApo_df <- data.frame(exprs(gs3DeadApo[[lec]]))

gs3_DeadApo_df <- mutate(gs3_DeadApo_df, Area_ratio = Pacific.Blue.A/FSC.A,
Height_ratio = Pacific.Blue.H/FSC.H)
```

```
gs3_DeadApo_df <- gs3_DeadApo_df[, -16]

gs1_Alive_df <- data.frame(exprs(gs1Alive[[lec]]))

gs1_Alive_df <- mutate(gs1_Alive_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio =
Pacific.Blue.H/FSC.H)

gs1_Alive_df <- gs1_Alive_df[, -16]

gs2_Alive_df <- data.frame(exprs(gs2Alive[[lec]]))

gs2_Alive_df <- mutate(gs2_Alive_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio =
Pacific.Blue.H/FSC.H)

gs2_Alive_df <- gs2_Alive_df[, -16]

gs3_Alive_df <- data.frame(exprs(gs3Alive[[lec]]))

gs3_Alive_df <- mutate(gs3_Alive_df, Area_ratio = Pacific.Blue.A/FSC.A, Height_ratio =
Pacific.Blue.H/FSC.H)

gs3_Alive_df <- gs3_Alive_df[, -16]

#Organizing everything on a list

filter_list <- list(list(gs1_Alive_1x_df, gs2_Alive_1x_df, gs3_Alive_1x_df),
list(gs1_Alive_int_df, gs2_Alive_int_df, gs3_Alive_int_df), list(gs1_Alive_2x_df,
gs2_Alive_2x_df, gs3_Alive_2x_df), list(gs1_Dead_df, gs2_Dead_df, gs3_Dead_df),
list(gs1_Apo_df, gs2_Apo_df, gs3_Apo_df), list(gs1_DeadApo_df, gs2_DeadApo_df,
gs3_DeadApo_df))

#Equalizing the sizes of samples from the three repeated experiments

Alive_1x_list <- equalize_samplesize(filter_list[[1]][[1]], filter_list[[1]][[2]], filter_list[[1]][[3]])

Alive_Int_list <- equalize_samplesize(filter_list[[2]][[1]], filter_list[[2]][[2]], filter_list[[2]][[3]])

Alive_2x_list <- equalize_samplesize(filter_list[[3]][[1]], filter_list[[3]][[2]], filter_list[[3]][[3]])

Dead_list <- equalize_samplesize(filter_list[[4]][[1]], filter_list[[4]][[2]], filter_list[[4]][[3]])

Apo_list <- equalize_samplesize(filter_list[[5]][[1]], filter_list[[5]][[2]], filter_list[[5]][[3]])

DeadApo_list <- equalize_samplesize(filter_list[[6]][[1]], filter_list[[6]][[2]],
filter_list[[6]][[3]])

#Calculating the individual experiment means, sd's, and cv's

Alive_2x_mean_list <- col_means(Alive_2x_list, names_vector)

Alive_Int_mean_list <- col_means(Alive_Int_list, names_vector)

Alive_1x_mean_list <- col_means(Alive_1x_list, names_vector)

Dead_mean_list <- col_means(Dead_list, names_vector)
```

```
Apo_mean_list <- col_means(Apo_list, names_vector)
DeadApo_mean_list <- col_means(DeadApo_list, names_vector)
Alive_2x_sd_list <- col_sd(Alive_2x_list, names_vector)
Alive_Int_sd_list <- col_sd(Alive_Int_list, names_vector)
Alive_1x_sd_list <- col_sd(Alive_1x_list, names_vector)
Dead_sd_list <- col_sd(Dead_list, names_vector)
Apo_sd_list <- col_sd(Apo_list, names_vector)
DeadApo_sd_list <- col_sd(DeadApo_list, names_vector)

Alive_2x_cv_list <- col_cv(Alive_2x_list, names_vector)
Alive_Int_cv_list <- col_cv(Alive_Int_list, names_vector)
Alive_1x_cv_list <- col_cv(Alive_1x_list, names_vector)
Dead_cv_list <- col_cv(Dead_list, names_vector)
Apo_cv_list <- col_cv(Apo_list, names_vector)
DeadApo_cv_list <- col_cv(DeadApo_list, names_vector)
#Calculating the mean of the means of the three repeated experiments, mean of the three
standard deviations and the mean of the three cv's
Alive_2x_control_global_mean <- col_means2(Alive_2x_mean_list, names_vector)
Alive_Int_control_global_mean <- col_means2(Alive_Int_mean_list, names_vector)
Alive_1x_control_global_mean <- col_means2(Alive_1x_mean_list, names_vector)
Dead_control_global_mean <- col_means2(Dead_mean_list, names_vector)
Apo_control_global_mean <- col_means2(Apo_mean_list, names_vector)
DeadApo_control_global_mean <- col_means2(DeadApo_mean_list, names_vector)
Alive_2x_control_global_sd <- col_sd2(Alive_2x_sd_list, names_vector)
Alive_Int_control_global_sd <- col_sd2(Alive_Int_sd_list, names_vector)
Alive_1x_control_global_sd <- col_sd2(Alive_1x_sd_list, names_vector)
Dead_control_global_sd <- col_sd2(Dead_sd_list, names_vector)
Apo_control_global_sd <- col_sd2(Apo_sd_list, names_vector)
DeadApo_control_global_sd <- col_sd2(DeadApo_sd_list, names_vector)
```

```
Alive_2x_control_global_cv <- col_cv2(Alive_2x_cv_list, names_vector)
Alive_Int_control_global_cv <- col_cv2(Alive_Int_cv_list, names_vector)
Alive_1x_control_global_cv <- col_cv2(Alive_1x_cv_list, names_vector)
Dead_control_global_cv <- col_cv2(Dead_cv_list, names_vector)
Apo_control_global_cv <- col_cv2(Apo_cv_list, names_vector)
DeadApo_control_global_cv <- col_cv2(DeadApo_cv_list, names_vector)
#Calculating the viability for the lectin file
gs1_viability <- (nrow(gs1_Alive_df) / (nrow(gs1_Alive_df) + nrow(gs1_DeadApo_df))) * 100
gs2_viability <- (nrow(gs2_Alive_df) / (nrow(gs2_Alive_df) + nrow(gs2_DeadApo_df))) * 100
gs3_viability <- (nrow(gs3_Alive_df) / (nrow(gs3_Alive_df) + nrow(gs3_DeadApo_df))) * 100
viability_mean <- mean(c(gs1_viability, gs2_viability, gs3_viability))
viability_sd <- sd(c(gs1_viability, gs2_viability, gs3_viability))
#Organizing the means, sd's and cv's on a list so it can be used in a loop
global_parameters_list <- list(list(Alive_2x_control_global_mean,
Alive_2x_control_global_sd, Alive_2x_control_global_cv),
list(Alive_Int_control_global_mean, Alive_Int_control_global_sd,
Alive_Int_control_global_cv), list(Alive_1x_control_global_mean,
Alive_1x_control_global_sd, Alive_1x_control_global_cv), list(Dead_control_global_mean,
Dead_control_global_sd,Dead_control_global_cv), list(Apo_control_global_mean,
Apo_control_global_sd, Apo_control_global_cv), list(DeadApo_control_global_mean,
DeadApo_control_global_sd,DeadApo_control_global_cv ))
equalized_filters_list <- list(Alive_2x_list, Alive_Int_list, Alive_1x_list, Dead_list, Apo_list,
DeadApo_list)
channel_names <- as.data.frame(matrix(names_vector, nrow = 17, ncol = 1),
stringsAsFactors = FALSE)
colnames(channel_names) <- c("Channels")
for (fn in 1:6) {
  for(fn2 in 1:3) {
    if (fn2 == 1) {
      control_table <- stack(global_parameters_list[[fn]][[fn2]])
      colnames(control_table) <- c("Mean","Channels2")
      control_table <- cbind(channel_names, control_table)
      control_table <- control_table[, -c(ncol(control_table))]
```

```
    } else if (fn2 == 2) {
      control_table1 <- stack(global_parameters_list[[fn]][[fn2]])
      colnames(control_table1) <- c("Mean SD","Channels2")
      control_table <- cbind(control_table, control_table1)
      control_table <- control_table[, -c(ncol(control_table))]
    } else {
      control_table2 <- stack(global_parameters_list[[fn]][[fn2]])
      colnames(control_table2) <- c("CV(%)","Channels2")
      control_table <- cbind(control_table, control_table2)
      control_table <- control_table[, -c(ncol(control_table))]
    }
  }
filter_label <- as.data.frame(matrix(c(rep(c(filter_names_vec[fn]), times =
nrow(control_table))), nrow = nrow(control_table), ncol = 1), stringsAsFactors = FALSE)
  colnames(filter_label) <- c("Subpopulation")
  sample_size_byfilter <- as.data.frame(matrix(rep(c(nrow(equalized_filters_list[[fn]][[1]])),
times = nrow(control_table)), nrow = nrow(control_table), ncol = 1))
  colnames(sample_size_byfilter) <- c("Sample Size")
  filter_loop_list[[fn]] <- cbind(control_table, filter_label, sample_size_byfilter)
}
viability_mean_matrix <- matrix(c(rep(viability_mean, times = 102)), nrow = 102, ncol = 1)
colnames(viability_mean_matrix) <- c("Viability(%)")
viability_sd_matrix <- matrix(c(rep(viability_sd, times = 102)), nrow = 102, ncol = 1)
colnames(viability_sd_matrix) <- c("Viability SD(%)")

#Add lectin name and viability info
lectin_names_vec <- c("AAL", "AAL-2", "MAL II", "PNA", "WGA", "LEC A", "LEC B")
lectin_name_matrix <- as.data.frame(matrix(c(rep(c(lectin_names_vec[lec - 10]), times =
102)), nrow = 102, ncol = 1), stringsAsFactors = FALSE)
colnames(lectin_name_matrix) <- c("Lectin")
```

```
546 lectin_loop_list[[lec - 10]] <- cbind(viability_mean_matrix,viability_sd_matrix,
547 lectin_name_matrix)

548 rbinding_filter_loop_list [[lec - 10]] <- rbind(filter_loop_list[[1]],
549 filter_loop_list[[2]],filter_loop_list[[3]],filter_loop_list[[4]],filter_loop_list[[5]],filter_loop_list
550 [[6]])

551 filter_loop_list <- vector(mode = "list")

552 }

553 table <-
554 rbind(cbind(rbinding_filter_loop_list[[1]],lectin_loop_list[[1]]),cbind(rbinding_filter_loop_list
555 [[2]],lectin_loop_list[[2]]),cbind(rbinding_filter_loop_list[[3]],lectin_loop_list[[3]]),
556 cbind(rbinding_filter_loop_list[[4]],lectin_loop_list[[4]]),cbind(rbinding_filter_loop_list[[5]],l
557 ectin_loop_list[[5]]), cbind(rbinding_filter_loop_list[[6]],lectin_loop_list[[6]]),

558 cbind(rbinding_filter_loop_list[[7]],lectin_loop_list[[7]]))

559 #Add Sample Type name

560 sample_name_matrix <- as.data.frame(matrix(c(rep(sample_type, times = 7 * 102)), nrow =
561 7 * 102, ncol = 1), stringsAsFactors = FALSE)

562 colnames(sample_name_matrix) <- c("Sample Type")

563 baseline_df <-  as.data.frame(cbind(sample_name_matrix, table), stringsAsFactors = FALSE)

564 baseline_df

565 }

566 #End of table_summary()

567 #Creating a function for the F-test

568 #Function argument is a data set containing info by filter and by lectin(sample/file). They
569 are both variables of the equalized sample sizes

570 F_T_Power_test <- function (baseline_df, sample_df, treatment_vec) {

571 #Lists for the construction of the data frame containing all the information processed
572 across the 6 filters(subpopulations) of each file(lectin)

573 f_test_list <- vector(mode = "list")

574 t_test_list <- vector(mode = "list")

575 power_test_list <- vector(mode = "list")

576 #Constructing a loop to go through the lectins(file 11 to 17)

577 for (lec in 1:7) {

578 #Constructing a for a loop to go through the subpopulations (filter)
```

```
  if (lec == 1) {
    r_min <- 1
    r_max <- 102
  } else if (lec == 2) {
    r_min <- 102 + 1
    r_max <- 102 * lec
  } else if (lec == 3) {
    r_min <- (102 * (lec-1)) +1
    r_max <- 102 * lec
  } else if (lec == 4) {
    r_min <- (102 * (lec-1)) +1
    r_max <- 102 * lec
  } else if (lec == 5) {
    r_min <- (102 * (lec-1)) +1
    r_max <- 102 * lec

  } else if (lec == 6) {
    r_min <- (102 * (lec-1)) +1
    r_max <- 102 * lec
  } else {
    r_min <- (102 * (lec-1)) +1
    r_max <- 102 * lec
  }
  for (r in r_min:r_max) {
control_size <- baseline_df[r, "Sample Size"]
sample_size <-  sample_df[r, "Sample Size"]
#Calcultating the degrees of freedom
df1 <- control_size - 1
df2 <- sample_size – 1
```

```
#Calculation of the f statistic

f_statistic <- ((baseline_df[r,"Mean SD"])^2) / ((sample_df[r,"Mean SD"])^2)

#Computation of the p-value. lower.tail is set to TRUE as the f test is two-sided (not need
to select the largest variance for the numerator for the computation of the f statistic)

fp_value <- pf(f_statistic, df1 = df1, df2 = df2, lower.tail = TRUE)

#Evaluation of the level of significance and decision making on equality or inequality of
variances

if ((fp_value >= 0.01) & (fp_value < 0.05)) {

f_list <- list(sample_df[r, "Mean"], sample_df[r, "Mean SD"], sample_size, fp_value,
"significant", "unequal variances")

} else if ((fp_value >= 0.001) & (fp_value <0.01)) {

f_list <- list(sample_df[r, "Mean"], sample_df[r, "Mean SD"], sample_size, fp_value, "highly
significant", "unequal variances")

} else if (fp_value < 0.001) {

f_list <- list(sample_df[r, "Mean"], sample_df[r, "Mean SD"], sample_size, fp_value, "very
highly significant", "unequal variances")

} else if ((fp_value >= 0.05) & (fp_value < 0.10)) {

f_list<- list(sample_df[r, "Mean"], sample_df[r, "Mean SD"], sample_size, fp_value, "trend
toward significance/not significant", "equal variances")

} else {

   f_list <- list(sample_df[r, "Mean"], sample_df[r, "Mean SD"], sample_size, fp_value, "not
significant", "equal variances")

}

#Construction of a variable containing the results of the F test evaluation

f_list_df <- as.data.frame(f_list, stringsAsFactors = FALSE)

colnames(f_list_df) <- c("Sample Mean", "Mean SD", "Sample Size", "F p-value", "Level of
siginificance", "F-test conclusion")

f_test_list[[r]] <- f_list_df

if (fp_value >= 0.05) {#Calculation of the t statistic for equal variances

s <- sqrt((df1*baseline_df[r,"Mean SD"]^2 + df2*sample_df[r,"Mean SD"]^2) / (control_size
+ sample_size - 2))

t_statistic <- (baseline_df[r,"Mean"] - sample_df[r,"Mean"]) / (s * sqrt(1/control_size +
1/sample_size))
```

```
dfree = control_size + sample_size - 2
tp_value <- pt(t_statistic, df = dfree, lower.tail = TRUE)
#Evaluation of the level of significance and decision making on the significance of the
difference between the two samples
if (tp_value >= 0.01 & tp_value <0.05) {
t_list <- list(tp_value, "significant")
}  else if (tp_value >= 0.001 & tp_value <0.01) {
  t_list <- list(tp_value, "highly significant")
} else if (tp_value < 0.001) {
  t_list<- list(tp_value, "very highly significant")
} else if (tp_value >= 0.05 & tp_value < 0.10) {
  t_list <- list(tp_value, "trend toward significance")
} else {
    t_list <- list(tp_value, "not significant")
}
} else {#Calculation of the t statistic for unequal variances
#T-Test for unequal variances (Satterthwaite's Method)
den <- sqrt(baseline_df[r, "Mean SD"]^2/control_size + sample_df[r, "Mean
SD"]^2/sample_size)
t_statistic <- (baseline_df[r, "Mean"] - sample_df[r, "Mean"])/ den

df_numerator <- (baseline_df[r, "Mean SD"]^2/control_size + sample_df[r, "Mean
SD"]^2/sample_size)^2
df_denominator <- (baseline_df[r, "Mean SD"]^2/control_size)^2/(control_size - 1) +
(sample_df[r, "Mean SD"]^2/sample_size )^2/(sample_size - 1)
dfree = df_numerator/df_denominator
tp_value <- pt(t_statistic, df = dfree, lower.tail = TRUE)
#Evaluation of the level of significance and decision making on the significance of the
difference between the two samples
if (tp_value >= 0.01 & tp_value <0.05) {
t_list <- list(tp_value, "significant")
```

```
672   }  else if (tp_value >= 0.001 & tp_value <0.01) {
673     t_list <- list(tp_value, "highly significant")
674   } else if (tp_value < 0.001) {
675     t_list<- list(tp_value, "very highly significant")
676   } else if (tp_value >= 0.05 & tp_value < 0.10) {
677     t_list <- list(tp_value, "trend toward significance")
678   } else {
679       t_list <- list(tp_value, "not significant")
680   }
681   #Adding the results of the T test evaluation to the variable created to store the data
682   }
683   t_list_df <- as.data.frame(t_list, stringsAsFactors = FALSE)
684   colnames(t_list_df) <- c("T p-value", "Level of siginificance")
685   t_test_list[[r]] <- t_list_df
686   #Compute Power Analysis before leaving this loop
687   c <- qnorm(0.975)
688   k <- sample_size / control_size # equal to n2/n1, that is, sample_size = n2 and control_size =
689   n1
690   se <- sqrt((baseline_df[r, "Mean SD"]^2)^2 + (sample_df[r, "Mean SD"]^2)^2 / k)
691   delta <- abs(baseline_df[r,"Mean"] - sample_df[r,"Mean"])
692   power <- pnorm(- c + (sqrt(control_size)*delta)/se)
693   power_df <- as.data.frame(c(power), stringsAsFactors = FALSE)
694   colnames(power_df) <- c("Power")
695   power_test_list[[r]] <- power_df
696   }
697   #r_min and r_max foor loop
698   }
699   #lectin for loop
700   for (i in 1:714) {
```

```
  if (i == 1 | i == 2) {

  f_test_complete_df <- as.data.frame(rbind(f_test_list[[1]], f_test_list[[2]]), stringsAsFactors
= FALSE)

  t_test_complete_df  <- as.data.frame(rbind(t_test_list[[1]], t_test_list[[2]]),
stringsAsFactors = FALSE)

  power_test_complete_df  <- as.data.frame(rbind(power_test_list[[1]],
power_test_list[[2]]), stringsAsFactors = FALSE)

  } else {

  f_test_complete_df <- as.data.frame(rbind(f_test_complete_df, f_test_list[[i]]),
stringsAsFactors = FALSE)

  t_test_complete_df  <- as.data.frame(rbind(t_test_complete_df, t_test_list[[i]]),
stringsAsFactors = FALSE)

  power_test_complete_df  <- as.data.frame(rbind(power_test_complete_df,
power_test_list[[i]]), stringsAsFactors = FALSE)

    }

}

f_t_power_complete_df <- as.data.frame(cbind(f_test_complete_df, t_test_complete_df,
power_test_complete_df))

treatment_matrix <- as.data.frame(matrix(c(rep(treatment_vec, times = 7 * 102)), nrow = 7
* 102, ncol = 1), stringsAsFactors = FALSE)

colnames(treatment_matrix) <- c("Treatment")

#Add the remaining lables from the baseline or sample_df

baseline_channels <- as.data.frame(baseline_df[, "Channels"], stringsAsFactors = FALSE)

baseline_subpopulation <- as.data.frame(baseline_df[, "Subpopulation"], stringsAsFactors =
FALSE)

baseline_lectin <- as.data.frame(baseline_df[, "Lectin"], stringsAsFactors = FALSE)

treat_channels_Subpopulation_Lectin_f_t_power_complete_df <-
as.data.frame(c(treatment_matrix, baseline_channels, baseline_subpopulation,
baseline_lectin, f_t_power_complete_df), stringsAsFactors = FALSE)

colnames(treat_channels_Subpopulation_Lectin_f_t_power_complete_df) <- c("Treatment",
"Channels", "Subpopulation", "Lectin", "Mean", "SD", "Sample Size", "F p-value", "F
significance","F test conclusion", "T p-value","T test significance", "Power")

treat_channels_Subpopulation_Lectin_f_t_power_complete_df

}
```

```
#End of  F_T_Power_test()

#Function to create a data frame containing the data from an experiment (triplicates
combined) for descriptive analysis only

table_descriptive <- function(x, y, z, sample_type) {

#Extracting flow data in variables

gs1Alive1x  <- getData(gs1,"/NonDebris/singletsa/singletsb/Alive/1xcells")

gs1AliveInt  <- getData(gs1,"/NonDebris/singletsa/singletsb/Alive/int")

gs1Alive2x  <- getData(gs1,"/NonDebris/singletsa/singletsb/Alive/2xcells")

gs1Dead  <- getData(gs1,"/NonDebris/singletsa/singletsb/Dead")

gs1Apo <- getData(gs1, "/NonDebris/singletsa/singletsb/Apoptotic")

gs1DeadApo <- getData(gs1, "/NonDebris/singletsa/singletsb/DeadApo")

gs1Alive <- getData(gs1,"/NonDebris/singletsa/singletsb/Alive")

gs2Alive1x  <- getData(gs2,"/NonDebris/singletsa/singletsb/Alive/1xcells")

gs2AliveInt  <- getData(gs2,"/NonDebris/singletsa/singletsb/Alive/int")

gs2Alive2x  <- getData(gs2,"/NonDebris/singletsa/singletsb/Alive/2xcells")

gs2Dead  <- getData(gs2,"/NonDebris/singletsa/singletsb/Dead")

gs2Apo<- getData(gs2, "/NonDebris/singletsa/singletsb/Apoptotic")

gs2DeadApo <- getData(gs2, "/NonDebris/singletsa/singletsb/DeadApo")

gs2Alive <- getData(gs2,"/NonDebris/singletsa/singletsb/Alive")

gs3Alive1x  <- getData(gs3,"/NonDebris/singletsa/singletsb/Alive/1xcells")

gs3AliveInt  <- getData(gs3,"/NonDebris/singletsa/singletsb/Alive/int")

gs3Alive2x  <- getData(gs3,"/NonDebris/singletsa/singletsb/Alive/2xcells")

gs3Dead  <- getData(gs3,"/NonDebris/singletsa/singletsb/Dead")

gs3Apo<- getData(gs3, "/NonDebris/singletsa/singletsb/Apoptotic")

gs3DeadApo <- getData(gs3, "/NonDebris/singletsa/singletsb/DeadApo")

gs3Alive <- getData(gs3,"/NonDebris/singletsa/singletsb/Alive")

#List for the contruction of the data frame containing all the information processed across
each file(lectin)

lectin_loop_list <- vector(mode = "list")

#Vector to label the flow cytometry channels, lectins and subpopulations identified
```

```
names_vector <- c("FSC_A", "FSC_H", "FSC_W", "SSC_A", "SSC_H", "SSC_W", "7AAD_A",
"7AAD_H", "7AAD_W", "DRAQ5_A", "DRAQ5_H", "DRAQ5_W", "LECTIN_A", "LECTIN_H",
"LECTIN_W")
lectin_names_vec <- c("AAL", "AAL-2", "MAL II", "PNA", "WGA", "LEC A", "LEC B")
#Loop to iterate though the lectin files (from 5th to 11th file)
for (lec in 11:17) {
#Extracting the data in a data frame format
gs1_Alive_1x_df <- data.frame(exprs(gs1Alive1x[[lec]]))
gs2_Alive_1x_df <- data.frame(exprs(gs2Alive1x[[lec]]))
gs3_Alive_1x_df <- data.frame(exprs(gs3Alive1x[[lec]]))
gs1_Alive_int_df <- data.frame(exprs(gs1AliveInt[[lec]]))
gs2_Alive_int_df <- data.frame(exprs(gs2AliveInt[[lec]]))
gs3_Alive_int_df <- data.frame(exprs(gs3AliveInt[[lec]]))

gs1_Alive_2x_df <- data.frame(exprs(gs1Alive2x[[lec]]))
gs2_Alive_2x_df <- data.frame(exprs(gs2Alive2x[[lec]]))
gs3_Alive_2x_df <- data.frame(exprs(gs3Alive2x[[lec]]))
gs1_Dead_df <- data.frame(exprs(gs1Dead[[lec]]))
gs2_Dead_df <- data.frame(exprs(gs2Dead[[lec]]))
gs3_Dead_df <- data.frame(exprs(gs3Dead[[lec]]))
gs1_Apo_df <- data.frame(exprs(gs1Apo[[lec]]))
gs2_Apo_df <- data.frame(exprs(gs2Apo[[lec]]))
gs3_Apo_df <- data.frame(exprs(gs3Apo[[lec]]))
gs1_DeadApo_df <- data.frame(exprs(gs1DeadApo[[lec]]))
gs2_DeadApo_df <- data.frame(exprs(gs2DeadApo[[lec]]))
gs3_DeadApo_df <- data.frame(exprs(gs3DeadApo[[lec]]))
gs1_Alive_df <- data.frame(exprs(gs1Alive[[lec]]))
gs2_Alive_df <- data.frame(exprs(gs2Alive[[lec]]))
gs3_Alive_df <- data.frame(exprs(gs3Alive[[lec]]))

```

```
#Organizing everything on a list

filter_list <- list(list(gs1_Alive_1x_df, gs2_Alive_1x_df, gs3_Alive_1x_df),
list(gs1_Alive_int_df, gs2_Alive_int_df, gs3_Alive_int_df), list(gs1_Alive_2x_df,
gs2_Alive_2x_df, gs3_Alive_2x_df), list(gs1_Dead_df, gs2_Dead_df, gs3_Dead_df),
list(gs1_Apo_df, gs2_Apo_df, gs3_Apo_df), list(gs1_DeadApo_df, gs2_DeadApo_df,
gs3_DeadApo_df))

#Equalizing the sizes of sample from the repeated experiments (3 times)

Alive_1x_list <- equalize_samplesize(filter_list[[1]][[1]], filter_list[[1]][[2]], filter_list[[1]][[3]])

Alive_Int_list <- equalize_samplesize(filter_list[[2]][[1]], filter_list[[2]][[2]], filter_list[[2]][[3]])

Alive_2x_list <- equalize_samplesize(filter_list[[3]][[1]], filter_list[[3]][[2]], filter_list[[3]][[3]])

Dead_list <- equalize_samplesize(filter_list[[4]][[1]], filter_list[[4]][[2]], filter_list[[4]][[3]])

Apo_list <- equalize_samplesize(filter_list[[5]][[1]], filter_list[[5]][[2]], filter_list[[5]][[3]])

DeadApo_list <- equalize_samplesize(filter_list[[6]][[1]], filter_list[[6]][[2]],
filter_list[[6]][[3]])

Alive_1x_df <- rbind(Alive_1x_list[[1]], Alive_1x_list[[2]], Alive_1x_list[[3]])

Alive_1x_df <- Alive_1x_df [, -16]

colnames(Alive_1x_df) <- names_vector

filter_label <- as.data.frame(matrix(c(rep(c("Go/G1"), times = nrow(Alive_1x_df))), nrow =
nrow(Alive_1x_df), ncol = 1), stringsAsFactors = FALSE)

colnames(filter_label) <- c("Subpopulation")

Alive_Int_df <- rbind(Alive_Int_list[[1]], Alive_Int_list[[2]], Alive_Int_list[[3]])

Alive_Int_df <- Alive_Int_df [, -16]

colnames(Alive_Int_df) <- names_vector

filter_label_temp <- as.data.frame(matrix(c(rep(c("S"), times = nrow(Alive_Int_df))), nrow =
nrow(Alive_Int_df), ncol = 1), stringsAsFactors = FALSE)

colnames(filter_label_temp) <- c("Subpopulation")

filter_label <- rbind(filter_label, filter_label_temp)

Alive_2x_df <- rbind(Alive_2x_list[[1]], Alive_2x_list[[2]], Alive_2x_list[[3]])

Alive_2x_df <- Alive_2x_df [, -16]

colnames(Alive_2x_df) <- names_vector

filter_label_temp <- as.data.frame(matrix(c(rep(c("G2/M"), times = nrow(Alive_2x_df))),
nrow = nrow(Alive_2x_df), ncol = 1), stringsAsFactors = FALSE)
```

```
827 colnames(filter_label_temp) <- c("Subpopulation")
828 filter_label <- rbind(filter_label, filter_label_temp)
829 Dead_df <- rbind(Dead_list[[1]], Dead_list[[2]], Dead_list[[3]])
830 Dead_df <- Dead_df [, -16]
831 colnames(Dead_df) <- names_vector
832 filter_label_temp <- as.data.frame(matrix(c(rep(c("Dead"), times = nrow(Dead_df))), nrow =
833 nrow(Dead_df), ncol = 1), stringsAsFactors = FALSE)
834 colnames(filter_label_temp) <- c("Subpopulation")
835 filter_label <- rbind(filter_label, filter_label_temp)
836 Apo_df <- rbind(Apo_list[[1]], Apo_list[[2]], Apo_list[[3]])
837 Apo_df <- Apo_df [, -16]
838 colnames(Apo_df) <- names_vector
839 filter_label_temp <- as.data.frame(matrix(c(rep(c("Apoptotic"), times = nrow(Apo_df))),
840 nrow = nrow(Apo_df), ncol = 1), stringsAsFactors = FALSE)
841 colnames(filter_label_temp) <- c("Subpopulation")
842 filter_label <- rbind(filter_label, filter_label_temp)
843 DeadApo_df <- rbind(DeadApo_list[[1]], DeadApo_list[[2]], DeadApo_list[[3]])
844 DeadApo_df <- DeadApo_df [, -16]
845 colnames(DeadApo_df) <- names_vector
846 filter_label_temp <- as.data.frame(matrix(c(rep(c("Dead + Apoptotic"), times =
847 nrow(DeadApo_df))), nrow = nrow(DeadApo_df), ncol = 1), stringsAsFactors = FALSE)
848 colnames(filter_label_temp) <- c("Subpopulation")
849 filter_label <- rbind(filter_label, filter_label_temp)
850 lectin_label <- as.data.frame(matrix(c(rep(c(lectin_names_vec[lec - 10]), times =
851 nrow(filter_label))), nrow = nrow(filter_label), ncol = 1), stringsAsFactors = FALSE)
852 colnames(lectin_label) <- c("Lectin")
853 lectin_loop_list[[lec - 10]] <- cbind(rbind(Alive_1x_df, Alive_Int_df, Alive_2x_df, Dead_df,
854 Apo_df, DeadApo_df), cbind(filter_label, lectin_label))
855 }
856 lectins_df <- rbind(lectin_loop_list[[1]],lectin_loop_list[[2]], lectin_loop_list[[3]],
857 lectin_loop_list[[4]], lectin_loop_list[[5]], lectin_loop_list[[6]], lectin_loop_list[[7]])
```

```
#Add Sample Type name

sample_name_matrix <- as.data.frame(matrix(c(rep(sample_type, times =
nrow(lectins_df))), nrow = nrow(lectins_df), ncol = 1), stringsAsFactors = FALSE)

colnames(sample_name_matrix) <- c("Sample_Type")

global_df <-  as.data.frame(cbind(sample_name_matrix, lectins_df), stringsAsFactors =
FALSE)

global_df <- mutate(global_df, Area_den = LECTIN_A/FSC_A, Height_den = LECTIN_H/FSC_H)

global_df

}

#End of the function table_descriptive()

#Creating a function to combine data from descriptive analysis (table_descriptive()) and
statistical significance levels ratios_stats_channel_choice()

#Example: FSC_W_Subp_GoG1_df <- table_manipulation(media_global_descriptive_df,
media_global_F_T_df, c("FSC_W"), c("FSC-W"), c("Go/G1"), c("Media"))

table_manipulation <- function(df1, df2, channel1, channel2, Subpop, variable)   {

lectin_names_vec <- c("AAL", "AAL-2", "MAL II", "PNA", "WGA", "LEC A", "LEC B")

lectin_loop_list <- vector(mode = "list")

sample_loop_list <- vector(mode = "list")

if (variable == c("Media")) {

Sample_Type_vec <- c("f", "e", "g", "c", "b", "a")

for (i in 1:6) {

for (lec in 1:7) {

table_1 <- select(df1, channel1, Subpopulation, Lectin, Sample_Type) %>% filter(Lectin ==
lectin_names_vec[lec], Sample_Type == Sample_Type_vec[i], Subpopulation == Subpop)

table_2 <- select(df2, Sample_Type, Channels, Subpopulation, Lectin, T_test_significance,
Power) %>% filter(Lectin == lectin_names_vec[lec], Sample_Type == Sample_Type_vec[i],
Channels == channel2, Subpopulation == Subpop) %>% select (T_test_significance)

T_test_matrix <- as.data.frame(matrix(c(rep(table_2[1,1], times = nrow(table_1))), nrow =
nrow(table_1), ncol = 1), stringsAsFactors = FALSE)

colnames(T_test_matrix) <- c("T_test_significance")

table_1_2 <- cbind(table_1, T_test_matrix)

lectin_loop_list[[lec]] <- table_1_2
```

```
890           }
891     sample_loop_list[[i]]  <- rbind(lectin_loop_list[[1]], lectin_loop_list[[2]], lectin_loop_list[[3]],
892     lectin_loop_list[[4]], lectin_loop_list[[5]], lectin_loop_list[[6]], lectin_loop_list[[7]])
893        }
894     global_table <- rbind(sample_loop_list[[1]], sample_loop_list[[2]], sample_loop_list[[3]],
895     sample_loop_list[[4]], sample_loop_list[[5]], sample_loop_list[[6]])
896        } else if (variable == c("Temp")) {
897     Sample_Type_vec <- c("32", "33", "34", "35", "36", "38", "39", "40")
898     for (i in 1:10) {
899     for (lec in 1:7) {
900     table_1 <- select(df1, channel1, Subpopulation, Lectin, Sample_Type) %>% filter(Lectin ==
901     lectin_names_vec[lec], Sample_Type == Sample_Type_vec[i], Subpopulation == Subpop)
902     table_2  <- select(df2, Sample_Type, Channels, Subpopulation, Lectin, T_test_significance,
903     Power) %>% filter(Lectin == lectin_names_vec[lec], Sample_Type == Sample_Type_vec[i],
904     Channels == channel2, Subpopulation == Subpop) %>% select (T_test_significance)
905     T_test_matrix <- as.data.frame(matrix(c(rep(table_2[1,1], times = nrow(table_1))), nrow =
906     nrow(table_1), ncol = 1), stringsAsFactors = FALSE)
907     colnames(T_test_matrix) <- c("T_test_significance")
908      table_1_2 <- cbind(table_1, T_test_matrix)
909      lectin_loop_list[[lec]] <- table_1_2
910           }
911      sample_loop_list[[i]]  <- rbind(lectin_loop_list[[1]], lectin_loop_list[[2]], lectin_loop_list[[3]],
912     lectin_loop_list[[4]], lectin_loop_list[[5]], lectin_loop_list[[6]], lectin_loop_list[[7]])
913        }
914     global_table <- rbind(sample_loop_list[[1]], sample_loop_list[[2]], sample_loop_list[[3]],
915     sample_loop_list[[4]], sample_loop_list[[5]], sample_loop_list[[6]], sample_loop_list[[7]],
916     sample_loop_list[[8]], sample_loop_list[[9]], sample_loop_list[[10]])
917       } else {
918     Sample_Type_vec <- c("a", "b", "c", "d", "f", "g", "h", "i", "j")
919        for (i in 1:9) {
920        for (lec in 1:7) {
921     table_1 <- select(df1, channel1, Subpopulation, Lectin, Sample_Type) %>% filter(Lectin ==
922     lectin_names_vec[lec], Sample_Type == Sample_Type_vec[i], Subpopulation == Subpop)
```

```
table_2 <- select(df2, Sample_Type, Channels, Subpopulation, Lectin, T_test_significance,
Power) %>% filter(Lectin == lectin_names_vec[lec], Sample_Type == Sample_Type_vec[i],
Channels == channel2, Subpopulation == Subpop) %>% select (T_test_significance)

 T_test_matrix <- as.data.frame(matrix(c(rep(table_2[1,1], times = nrow(table_1))), nrow =
nrow(table_1), ncol = 1), stringsAsFactors = FALSE)

colnames(T_test_matrix) <- c("T_test_significance")

table_1_2 <- cbind(table_1, T_test_matrix)

lectin_loop_list[[lec]] <- table_1_2

   }

sample_loop_list[[i]]  <- rbind(lectin_loop_list[[1]], lectin_loop_list[[2]], lectin_loop_list[[3]],
lectin_loop_list[[4]], lectin_loop_list[[5]], lectin_loop_list[[6]], lectin_loop_list[[7]])

 }

global_table <- rbind(sample_loop_list[[1]], sample_loop_list[[2]], sample_loop_list[[3]],
sample_loop_list[[4]], sample_loop_list[[5]], sample_loop_list[[6]], sample_loop_list[[7]],
sample_loop_list[[8]], sample_loop_list[[9]])

 }

 global_table

}

#End of function table_manipulation()

lectin_density_stats <- function (dataset) {

dataset <- filter(dataset, Channels %in% c("Area_ratio", "Height_ratio"), Subpopulation %in%
c("G2/M", "S", "Go/G1"))

f_test_list <- vector(mode = "list")

t_test_list <- vector(mode = "list")

power_test_list <- vector(mode = "list")

nr <- 1

for (lec in 1:7) {

 #Constructing a for a loop to go through the subpopulations (filter)

 if (lec == 1) {

  r_min <- 1

  #r_max <- 6

 } else if (lec == 2) {
```

```
955       r_min <- 6 + 1
956       #r_max <- 6 * lec
957     } else if (lec == 3) {
958       r_min <- (6 * (lec-1)) +1
959       #r_max <- 6 * lec
960     } else if (lec == 4) {
961       r_min <- (6 * (lec-1)) +1
962       #r_max <- 6 * lec
963     } else if (lec == 5) {
964       r_min <- (6 * (lec-1)) +1
965       #r_max <- 6 * lec
966     } else if (lec == 6) {
967       r_min <- (6 * (lec-1)) +1
968       #r_max <- 6 * lec
969     } else {
970       r_min <- (6 * (lec-1)) +1
971       #r_max <- 6 * lec
972     }
973   for (i in 1:6) {
974     if(i == 1 | i == 2 | i == 3 | i == 4) {
975   size1 <- dataset[r_min, "Sample Size"]
976   size2 <-  dataset[r_min + 2, "Sample Size"]
977   #Calcultating the degrees of freedom
978   df1 <- size1 - 1
979   df2 <- size2 - 1
980   #Calculation of the f statistic
981   f_statistic <- ((dataset[r_min,"Mean SD"])^2) / ((dataset[r_min + 2,"Mean SD"])^2)
982   #Computation of the p-value. lower.tail is set to TRUE as the f test is two-sided
983   fp_value <- pf(f_statistic, df1 = df1, df2 = df2, lower.tail = TRUE)
```

```
#Evaluation of the level of significance and decision making on equality or inequality of variances
if ((fp_value >= 0.01) & (fp_value < 0.05)) {
f_list <- list(fp_value, "significant", "unequal variances")
} else if ((fp_value >= 0.001) & (fp_value <0.01)) {
f_list <- list(fp_value, "highly significant", "unequal variances")
} else if (fp_value < 0.001) {
f_list <- list(fp_value, "very highly significant", "unequal variances")
} else if ((fp_value >= 0.05) & (fp_value < 0.10)) {
f_list<- list(fp_value, "trend toward significance/not significant", "equal variances")
} else {
   f_list <- list(fp_value, "not significant", "equal variances")
}
#Construction of a variable containing the results of the F test evaluation
f_list_df <- as.data.frame(f_list, stringsAsFactors = FALSE)
colnames(f_list_df) <- c("F p-value", "Level of siginificance", "F-test conclusion")
f_test_list[[nr]] <- f_list_df
if (fp_value >= 0.05) {#Calculation of the t statistic for equal variances
s <- sqrt((df1*dataset[r_min,"Mean SD"]^2 + df2*dataset[r_min + 2,"Mean SD"]^2) / (size1 + size2 - 2))
t_statistic <- (dataset[r_min,"Mean"] - dataset[r_min + 2,"Mean"]) / (s * sqrt(1/size1 + 1/size2))
dfree = size1 + size2 - 2
tp_value <- pt(t_statistic, df = dfree, lower.tail = TRUE)
#Evaluation of the level of significance and decision making on the significance of the difference between the two samples
if (tp_value >= 0.01 & tp_value <0.05) {
t_list <- list(tp_value, "significant")
}  else if (tp_value >= 0.001 & tp_value <0.01) {
  t_list <- list(tp_value, "highly significant")
} else if (tp_value < 0.001) {
```

```
  t_list<- list(tp_value, "very highly significant")
} else if (tp_value >= 0.05 & tp_value < 0.10) {
  t_list <- list(tp_value, "trend toward significance")
} else {
    t_list <- list(tp_value, "not significant")
}
} else {#Calculation of the t statistic for unequal variances
#T-Test for unequal variances (Satterthwaite's Method)
den <- sqrt(dataset[r_min, "Mean SD"]^2/size1 + dataset[r_min + 2, "Mean SD"]^2/size2)
t_statistic <- (dataset[r_min, "Mean"] - dataset[r_min + 2, "Mean"])/ den
df_numerator <- (dataset[r_min, "Mean SD"]^2/size1 + dataset[r_min + 2, "Mean
SD"]^2/size2)^2
df_denominator <- (dataset[r_min, "Mean SD"]^2/size1)^2/(size1 - 1) + (dataset[r_min + 2,
"Mean SD"]^2/size2 )^2/(size2 - 1)
dfree = df_numerator/df_denominator
tp_value <- pt(t_statistic, df = dfree, lower.tail = TRUE)
#Evaluation of the level of significance and decision making on the significance of the
difference between the two samples
if (tp_value >= 0.01 & tp_value <0.05) {
t_list <- list(tp_value, "significant")
}  else if (tp_value >= 0.001 & tp_value <0.01) {
  t_list <- list(tp_value, "highly significant")
} else if (tp_value < 0.001) {
t_list<- list(tp_value, "very highly significant")
} else if (tp_value >= 0.05 & tp_value < 0.10) {
t_list <- list(tp_value, "trend toward significance")
} else {
t_list <- list(tp_value, "not significant")
}
#Adding the results of the T test evaluation to the variable created to store the data
```

```
}
t_list_df <- as.data.frame(t_list, stringsAsFactors = FALSE)
colnames(t_list_df) <- c("T p-value", "Level of siginificance")
t_test_list[[nr]] <- t_list_df
#Compute Power Analysis before leaving this loop
c <- qnorm(0.975)
k <- size2 / size1 # equal to n2/n1, that is, sample_size = n2 and control_size = n1
se <- sqrt((dataset[r_min, "Mean SD"]^2)^2 + (dataset[r_min + 2, "Mean SD"]^2)^2 / k)
delta <- abs(dataset[r_min,"Mean"] - dataset[r_min + 2,"Mean"])
power <- pnorm(- c + (sqrt(size1)*delta)/se)
power_df <- as.data.frame(c(power), stringsAsFactors = FALSE)
colnames(power_df) <- c("Power")
power_test_list[[nr]] <- power_df
#r_min and nr increment
r_min <- r_min + 1
nr <- nr + 1
  } else {

size1 <- dataset[r_min, "Sample Size"]
size2 <- dataset[r_min - 4, "Sample Size"]
#Calcultating the degrees of freedom
df1 <- size1 - 1
df2 <- size2 - 1
#Calculation of the f statistic
f_statistic <- ((dataset[r_min,"Mean SD"])^2) / ((dataset[r_min - 4,"Mean SD"])^2)
#Computation of the p-value. lower.tail is set to TRUE as the f test is two-sided
fp_value <- pf(f_statistic, df1 = df1, df2 = df2, lower.tail = TRUE)
#Evaluation of the level of significance and decision making on equality or inequality of
variances
```

```
if ((fp_value >= 0.01) & (fp_value < 0.05)) {
f_list <- list(fp_value, "significant", "unequal variances")
} else if ((fp_value >= 0.001) & (fp_value <0.01)) {
f_list <- list(fp_value, "highly significant", "unequal variances")
} else if (fp_value < 0.001) {
f_list <- list(fp_value, "very highly significant", "unequal variances")
} else if ((fp_value >= 0.05) & (fp_value < 0.10)) {
f_list<- list(fp_value, "trend toward significance/not significant", "equal variances")
} else {
   f_list <- list(fp_value, "not significant", "equal variances")
}
#Construction of a variable containing the results of the F test evaluation
f_list_df <- as.data.frame(f_list, stringsAsFactors = FALSE)
colnames(f_list_df) <- c("F p-value", "Level of siginificance", "F-test conclusion")
f_test_list[[nr]] <- f_list_df
if (fp_value >= 0.05) {#Calculation of the t statistic for equal variances
s <- sqrt((df1*dataset[r_min,"Mean SD"]^2 + df2*dataset[r_min - 4,"Mean SD"]^2) / (size1 +
size2 - 2))
t_statistic <- (dataset[r_min,"Mean"] - dataset[r_min - 4,"Mean"]) / (s * sqrt(1/size1 +
1/size2))
dfree = size1 + size2 - 2
tp_value <- pt(t_statistic, df = dfree, lower.tail = TRUE)
#Evaluation of the level of significance and decision making on the significance of the
difference between the two samples
if (tp_value >= 0.01 & tp_value <0.05) {
t_list <- list(tp_value, "significant")
}  else if (tp_value >= 0.001 & tp_value <0.01) {
t_list <- list(tp_value, "highly significant")
} else if (tp_value < 0.001) {
t_list<- list(tp_value, "very highly significant")
```

```
} else if (tp_value >= 0.05 & tp_value < 0.10) {
t_list <- list(tp_value, "trend toward significance")
} else {
t_list <- list(tp_value, "not significant")
}
} else {#Calculation of the t statistic for unequal variances
#T-Test for unequal variances (Satterthwaite's Method)
 den <- sqrt(dataset[r_min, "Mean SD"]^2/size1 + dataset[r_min - 4, "Mean SD"]^2/size2)
t_statistic <- (dataset[r_min, "Mean"] - dataset[r_min - 4, "Mean"])/ den
df_numerator <- (dataset[r_min, "Mean SD"]^2/size1 + dataset[r_min - 4, "Mean
SD"]^2/size2)^2
df_denominator <- (dataset[r_min, "Mean SD"]^2/size1)^2/(size1 - 1) + (dataset[r_min - 4,
"Mean SD"]^2/size2 )^2/(size2 - 1)
dfree = df_numerator/df_denominator
tp_value <- pt(t_statistic, df = dfree, lower.tail = TRUE)

#Evaluation of the level of significance and decision making on the significance of the
difference between the two samples
if (tp_value >= 0.01 & tp_value <0.05) {
t_list <- list(tp_value, "significant")
}  else if (tp_value >= 0.001 & tp_value <0.01) {
t_list <- list(tp_value, "highly significant")
} else if (tp_value < 0.001) {
t_list<- list(tp_value, "very highly significant")
} else if (tp_value >= 0.05 & tp_value < 0.10) {
t_list <- list(tp_value, "trend toward significance")
} else {
   t_list <- list(tp_value, "not significant")
}
#Adding the results of the T test evaluation to the variable created to store the data
```

```
1134    }
1135    t_list_df <- as.data.frame(t_list, stringsAsFactors = FALSE)
1136    colnames(t_list_df) <- c("T p-value", "Level of siginificance")
1137    t_test_list[[nr]] <- t_list_df
1138
1139    #Compute Power Analysis before leaving this loop
1140    c <- qnorm(0.975)
1141    k <- size2 / size1 # equal to n2/n1, that is, sample_size = n2 and control_size = n1
1142    se <- sqrt((dataset[r_min, "Mean SD"]^2)^2 + (dataset[r_min - 4, "Mean SD"]^2)^2 / k)
1143    delta <- abs(dataset[r_min,"Mean"] - dataset[r_min - 4,"Mean"])
1144    power <- pnorm(- c + (sqrt(size1)*delta)/se)
1145    power_df <- as.data.frame(c(power), stringsAsFactors = FALSE)
1146    colnames(power_df) <- c("Power")
1147    power_test_list[[nr]] <- power_df
1148    #r_min and nr increment
1149      r_min <- r_min + 1
1150      nr <- nr + 1
1151       }
1152     }
1153    }
1154    for (i in 1:42) {
1155     if (i == 1 | i == 2) {
1156     f_test_complete_df <- as.data.frame(rbind(f_test_list[[1]], f_test_list[[2]]), stringsAsFactors
1157    = FALSE)
1158     t_test_complete_df  <- as.data.frame(rbind(t_test_list[[1]], t_test_list[[2]]),
1159    stringsAsFactors = FALSE)
1160     power_test_complete_df  <- as.data.frame(rbind(power_test_list[[1]],
1161    power_test_list[[2]]), stringsAsFactors = FALSE)
1162     } else {
1163     f_test_complete_df <- as.data.frame(rbind(f_test_complete_df, f_test_list[[i]]),
1164    stringsAsFactors = FALSE)
```

```
1165    t_test_complete_df  <- as.data.frame(rbind(t_test_complete_df, t_test_list[[i]]),
1166  stringsAsFactors = FALSE)

1167    power_test_complete_df  <- as.data.frame(rbind(power_test_complete_df,
1168  power_test_list[[i]]), stringsAsFactors = FALSE)

1169      }

1170  }

1171  f_t_power_complete_df <- as.data.frame(cbind(f_test_complete_df, t_test_complete_df,
1172  power_test_complete_df))

1173  treatment_matrix <- as.data.frame(matrix(c(rep(dataset[1, "Sample Type"], times = 42)),
1174  nrow = 42, ncol = 1), stringsAsFactors = FALSE)

1175  colnames(treatment_matrix) <- c("Treatment")

1176  comparison_type_matrix <- as.data.frame(matrix(c("B", "B", "A", "A", "C", "C"), nrow = 6,
1177  ncol = 1), stringsAsFactors = FALSE)

1178  #Comp_type A = Go/G1 vs S, B = S vs G2/M, and C = Go/G1 vs G2/M

1179  comparison_type_matrix <- rbind(comparison_type_matrix, comparison_type_matrix,
1180  comparison_type_matrix, comparison_type_matrix, comparison_type_matrix,
1181  comparison_type_matrix, comparison_type_matrix)

1182  colnames(comparison_type_matrix) <- c("Comp_Type")

1183  #Add the remaining lables from the dataset

1184  baseline_channels <- as.data.frame(dataset[, "Channels"], stringsAsFactors = FALSE)

1185  baseline_lectin <- as.data.frame(dataset[, "Lectin"], stringsAsFactors = FALSE)

1186  treat_channels_compType_Lectin_f_t_power_complete_df <-
1187  as.data.frame(c(treatment_matrix, baseline_channels, baseline_lectin,
1188  f_t_power_complete_df, comparison_type_matrix), stringsAsFactors = FALSE)

1189  colnames(treat_channels_compType_Lectin_f_t_power_complete_df)  <- c("Treatment",
1190  "Channels", "Lectin","Fp_value", "F_sign","F_test_con", "Tp_value","T_test_sig", "Power",
1191  "Comp_type")

1192  treat_channels_compType_Lectin_f_t_power_complete_df

1193  }

1194  #End of lectin_density_stats() function

1195
```

8.2 Spent medium data treatment and generation of plots

Data obtained from cells subjected to the variation of spent medium levels are computed in this section. Pre built-in R functions and the functions created in the previous section are used here. The code for the generation of plots are demonstrated in this section as well.

```
#Algorithm to collect gated data of the Media Variation Experiments
setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/Baseline")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments
II/Baseline/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/0 day SM")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/0 day
SM/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
```

```
save_gs(gs3, path = file.path(wd, "gs3"))

setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/1 day SM")

wd <- getwd()

x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/1 day
SM/Compensation Controls - WGA")

flow_gating_list <- flow_gating(wd, x_WGA)

#flow_gating_list <- flow_gating(wd)

gs1 <- flow_gating_list[[1]]

gs2 <- flow_gating_list[[2]]

gs3 <- flow_gating_list[[3]]

save_gs(gs1, path = file.path(wd, "gs1"))

save_gs(gs2, path = file.path(wd, "gs2"))

save_gs(gs3, path = file.path(wd, "gs3"))

setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/2 day SM")

wd <- getwd()

x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/2 day
SM/Compensation Controls - WGA")

flow_gating_list <- flow_gating(wd, x_WGA)

#flow_gating_list <- flow_gating(wd)

gs1 <- flow_gating_list[[1]]

gs2 <- flow_gating_list[[2]]

gs3 <- flow_gating_list[[3]]

save_gs(gs1, path = file.path(wd, "gs1"))

save_gs(gs2, path = file.path(wd, "gs2"))

save_gs(gs3, path = file.path(wd, "gs3"))

setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/4 day SM")

wd <- getwd()

x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/4 day
SM/Compensation Controls - WGA")

flow_gating_list <- flow_gating(wd, x_WGA)
```

```
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/5 day SM")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/5 day
SM/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/6 day SM")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/6 day
SM/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
```

```
save_gs(gs3, path = file.path(wd, "gs3"))
```

Algorithm to retreive gated data and run the statistical analysis of the Media Variation Experiments

```{r}
setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/Baseline")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
media_baseline <- table_summary(gs1, gs2, gs3, c("d"))
media_baseline_descriptive <- table_descriptive(gs1, gs2, gs3, c("d"))
media_baseline_density <- lectin_density_stats(media_baseline)

setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/0 day SM")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
media_zero <- table_summary(gs1, gs2, gs3, c("g"))
media_zero_descriptive <- table_descriptive(gs1, gs2, gs3, c("g"))
base_zero <- F_T_Power_test(media_baseline, media_zero, c("g"))
media_zero_density <- lectin_density_stats(media_zero)
setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/1 day SM")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
media_one <- table_summary(gs1, gs2, gs3, c("f"))
```

```
media_one_descriptive <- table_descriptive(gs1, gs2, gs3, c("f"))
base_one <- F_T_Power_test(media_baseline, media_one, c("f"))
media_one_density <- lectin_density_stats(media_one)
setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/2 day SM")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
media_two <- table_summary(gs1, gs2, gs3, c("e"))
media_two_descriptive <- table_descriptive(gs1, gs2, gs3, c("e"))
base_two <- F_T_Power_test(media_baseline, media_two, c("e"))
media_two_density <- lectin_density_stats(media_two)
setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/4 day SM")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
media_four <- table_summary(gs1, gs2, gs3, c("c"))
media_four_descriptive <- table_descriptive(gs1, gs2, gs3, c("c"))
base_four <- F_T_Power_test(media_baseline, media_four, c("c"))
media_four_density <- lectin_density_stats(media_four)
setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/5 day SM")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
media_five <- table_summary(gs1, gs2, gs3, c("b"))
media_five_descriptive <- table_descriptive(gs1, gs2, gs3, c("b"))
base_five <- F_T_Power_test(media_baseline, media_five, c("b"))
```

```
media_five_density <- lectin_density_stats(media_five)

setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II/6 day SM")

wd <- getwd()

gs1 <- load_gs(file.path(wd, "gs1"))

gs2 <- load_gs(file.path(wd, "gs2"))

gs3 <- load_gs(file.path(wd, "gs3"))

media_six <- table_summary(gs1, gs2, gs3, c("a"))

media_six_descriptive <- table_descriptive(gs1, gs2, gs3, c("a"))

base_six <- F_T_Power_test(media_baseline, media_six, c("a"))

media_six_density <- lectin_density_stats(media_six)

media_global_descriptive_df <- rbind(media_baseline_descriptive, media_zero_descriptive,
media_one_descriptive, media_two_descriptive, media_four_descriptive,
media_five_descriptive, media_six_descriptive)

rm(media_baseline_descriptive, media_zero_descriptive, media_one_descriptive,
media_two_descriptive, media_four_descriptive, media_five_descriptive,
media_six_descriptive)

media_global_lectinvariation_df <- rbind(media_baseline, media_zero, media_one,
media_two, media_four, media_five, media_six)

rm(media_baseline, media_zero, media_one, media_two, media_four, media_five,
media_six)

media_global_F_T_df <- rbind(base_zero, base_one, base_two, base_four, base_five,
base_six)

rm(base_zero, base_one, base_two, base_four, base_five, base_six)

media_global_density_df <- rbind(media_baseline_density, media_zero_density,
media_two_density, media_four_density, media_five_density, media_six_density)

rm(media_baseline_density, media_zero_density, media_two_density, media_four_density,
media_five_density, media_six_density)

colnames(media_global_F_T_df) <- c("Sample_Type", "Channels", "Subpopulation", "Lectin",
"Mean", "SD", "Sample_Size","Fp_value", "F_significance", "F_test_conclusion", "Tp_value",
"T_test_significance", "Power")

colnames(media_global_lectinvariation_df) <- c("Sample_Type", "Channels", "Mean",
"Mean_SD", "CV_perc", "Subpopulation", "Sample_Size","Viability_perc",
"Viability_SD_perc", "Lectin")

```

Plotting pH

```{r}
library(readxl)
library(gridExtra)
**#Read in the excel spreadsheet into R**
setwd("~/Dropbox/PhD Project/PhD Project/Media Depletion Experiments II")
pH_media <- read_excel("pH.xlsx")
pH_media_df <- as.data.frame(pH_media, stringsAsFactors = FALSE)
```

```{r}
```

Viability and pH Plots

```{r}
**#Viability across nutrient variation (line plot of individual lectin curves)**
viability_plot <- mutate(media_global_lectinvariation_df, Sample_Type =
factor(Sample_Type, levels = c("6", "5", "4", "3", "2", "1", "0"))) %>%
  ggplot(aes(Sample_Type, Viability_perc)) +
  geom_smooth(aes(group = Lectin, color = Lectin), size = 1.5, se = FALSE) +
  scale_colour_manual(name = "Lectin",  values = c("#980043","#7a0177", "#08519c",
"#006d2c", "#7fcdbb",  "#ff7f00", "#993404")) +
   labs(x = "Level of Spent medium (Days)" , y = "Viability (%)", title = NULL ) +
scale_x_discrete(expand = c(0,0), breaks = c("6", "5", "4", "3", "2", "1", "0"), labels = c("-3", "-
2", "-1", "0", "+1", "+2", "+3")) +
  theme_classic() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
```

```
    legend.position = "right",
    #axis.text.x = element_text(size = 10, face = "bold", color = "black"),
    #axis.ticks.x = element_blank(),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90")
  )
#pH line plot across spent medium variation
ggplot(aes(Lectin_Concentration, Viability, group = 1)) +
  geom_point(aes(colour = Replicate), size = 3, alpha = 0.60) +
  geom_smooth(colour = "#993404", method = "lm", size = 1.5, formula = my.formula) +
  #stat_poly_eq(formula = my.formula, aes(label = paste(..eq.label.., ..rr.label.., sep =
"~~~")), parse = TRUE, label.x = 0.5, label.y = 0.2) +
   labs(x = expression(paste("Lectin concentration (", mu, "g/mL)")) , y = "Viability (%)", title =
NULL ) +
  scale_colour_brewer(palette = "Set1", name = "Replicate") +
pH_plot <- ggplot(pH_media_df, aes(Sample_Type, pH)) +
  geom_point(aes(colour = Replicate), size = 3, alpha = 0.60) +
  geom_smooth(size = 1.5) +
  labs(x = "Level of Spent medium (Days)" , y = "pH", title = NULL ) +
  scale_colour_brewer(palette = "Set1", name = "Replicate") +
    scale_x_continuous(expand = c(0, 0), breaks = c(-3, -2, -1, 0, 1, 2, 3), labels = c("-3", "-2", "-
1", "0", "+1", "+2", "+3")) +
  scale_y_continuous(expand = c(0,0)) +
  theme_classic() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
```

```
    legend.justification = "center",
    legend.position = "right",
    #axis.text.x = element_text(size = 10, face = "bold", color = "black"),
    #axis.ticks.x = element_blank(),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90")
  )
ggplot(media_global_lectinvariation_df, aes(Sample_Type, Viability_perc)) +
  geom_point(alpha = 0.0) +
 geom_smooth(data = viabilityPNA_df, aes(colour = "A"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 8), se = FALSE) +
  geom_smooth(data = viabilityAAL_df, aes(colour = "B"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 7), se = FALSE) +
  geom_smooth(data = viabilityMALII_df, aes(colour = "C"), method = "lm", size = 1.5,
formula = y ~ splines::bs(x, 8), se = FALSE) +
  geom_smooth(data = viabilityLECB_df, aes(colour = "D"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 8), se = FALSE) +
  geom_smooth(data = viabilityLECA_df, aes(colour = "E"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 8), se = FALSE) +
  geom_smooth(data = viabilityAAL2_df, aes(colour = "F"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 8), se = FALSE) +
  geom_smooth(data = viabilityWGA_df, aes(colour = "G"), method = "lm", size = 1.5,
formula = y ~ splines::bs(x, 6), se = FALSE) +
geom_vline(aes(xintercept = c(6.95)), color = "red", linetype = "dashed", size = 1) +
#geom_text(aes(x = 5.5, label = expression(paste("Lectin concentration level\n selected at
3.0", mu, "g/mL")) , y = 96), colour="red", angle = 0) +
  labs(x = expression(paste("Lectin concentration (", mu, "g/mL")) , y = "Viability (%)", title =
NULL ) +
  scale_colour_manual(name = "Lectin",  values = c("#ff7f00", "#980043", "#7fcdbb",
"#006d2c", "#08519c", "#7a0177", "#993404"), breaks = c("A", "B", "C", "D", "E", "F", "G"),
labels = c("PNA", "AAL", "MALII", "LECB", "LECA", "AAL-2", "WGA")) +
scale_x_discrete(expand = c(0,0)) +
  theme_classic() +
```

```
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "right",
    #axis.text.x = element_text(size = 10, face = "bold", color = "black"),
    #axis.ticks.x = element_blank(),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
    #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
    #legend.background = element_rect(fill = "grey90", colour = "grey90")
  )
```

Facetted plots with all lectins - Descriptive Analysis

```{r}
#FSC-A
#AAL, LECB, PNA, LECA, AAL-2, WGA, MAL II
filter(media_global_lectinvariation_df, Channels == "FSC-A", Subpopulation != "Dead + Apoptotic") %>%
mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S", "Go/G1"))) %>%
mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))) %>%
mutate(Sample_Type = factor(Sample_Type, levels = c("6", "5", "4", "3", "2", "1", "0"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
geom_point() +
```

```
geom_line(size = 1) +
theme_classic() +
labs(x = "Level of Spent medium (Days)" , y = "FSC-A (linear scale)", title = NULL ) +
scale_colour_brewer(palette = "Dark2", name = "Subpopulation") +
scale_x_discrete(expand = c(0,0), breaks = c("6", "5", "4", "3", "2", "1", "0"), labels = c("-3", "-
2", "-1", "0", "+1", "+2", "+3")) +
facet_grid(.~ Lectin) +
theme_bw() +
theme(
plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
legend.text = element_text(size = 15),
legend.title = element_text(size = 15, face = "bold"),
legend.box.background = element_blank(),
legend.justification = "center",
legend.position = "bottom",
#axis.text.x = element_text(angle = 45),
#axis.ticks.x = element_blank(),
axis.title = element_text(size = 15),
strip.text = element_text(size = 15),
strip.background = element_rect(fill = "grey90"),
panel.grid = element_blank(),
panel.spacing = unit(0.75, "lines")
#panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
#panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
#legend.background = element_rect(fill = "grey90", colour = "grey90")
  )
#SSC-A
#All populations
p1 <- filter(media_global_lectinvariation_df, Channels == "SSC-A", Subpopulation != "Dead +
Apoptotic") %>%
```

```
mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S",
"Go/G1"))) %>%

mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
"MAL II"))) %>%

mutate(Sample_Type = factor(Sample_Type, levels = c("6", "5", "4", "3", "2", "1", "0"))) %>%

ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +

geom_point() +

geom_line(size = 1) +

theme_classic() +

labs(x = NULL , y = "SSC-A (linear scale)", title = NULL ) +

scale_colour_brewer(palette = "Dark2", name = "Subpopulation", guide = FALSE) +

scale_x_discrete(expand = c(0,0), breaks = c("6", "5", "4", "3", "2", "1", "0"), labels = c("-3", "-
2", "-1", "0", "+1", "+2", "+3")) +

facet_grid(.~ Lectin) +

theme_bw() +

theme(

plot.title = element_text(face = "bold", size = 18, hjust = 0.5),

legend.text = element_text(size = 15),

legend.title = element_text(size = 15, face = "bold"),

legend.box.background = element_blank(),

legend.justification = "center",

legend.position = "bottom",

#axis.text.x = element_text(angle = 45),

#axis.ticks.x = element_blank(),

axis.title = element_text(size = 15),

strip.text = element_text(size = 15),

strip.background = element_rect(fill = "grey90"),

panel.grid = element_blank(),

panel.spacing = unit(0.75, "lines")

#panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
```

```
#panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
#legend.background = element_rect(fill = "grey90", colour = "grey90")
  )
#SSC-A
#DNA cycle populations
p2 <- filter(media_global_lectinvariation_df, Channels == "SSC-A", Subpopulation != "Dead + Apoptotic") %>%
mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S", "Go/G1"))) %>%
mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))) %>%
mutate(Sample_Type = factor(Sample_Type, levels = c("6", "5", "4", "3", "2", "1", "0"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
geom_point() +
geom_line(size = 1) +
theme_classic() +
labs(x = "Level of Spent medium (Days)" , y = "SSC-A (linear scale)", title = NULL ) +
scale_colour_brewer(palette = "Dark2", name = "Subpopulation") +
scale_x_discrete(expand = c(0,0), breaks = c("6", "5", "4", "3", "2", "1", "0"), labels = c("-3", "-2", "-1", "0", "+1", "+2", "+3")) +
scale_y_continuous(limits = c(20000, 55000)) +
facet_grid(.~ Lectin) +
theme_bw() +
theme(
plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
legend.text = element_text(size = 15),
legend.title = element_text(size = 15, face = "bold"),
legend.box.background = element_blank(),
legend.justification = "center",
legend.position = "bottom",
#axis.text.x = element_text(angle = 45),
```

```
#axis.ticks.x = element_blank(),
axis.title = element_text(size = 15),
strip.text = element_text(size = 15),
strip.background = element_rect(fill = "grey90"),
panel.grid = element_blank(),
panel.spacing = unit(0.75, "lines")
#panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
#panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
#legend.background = element_rect(fill = "grey90", colour = "grey90")
  )
grid.arrange(p1, p2, nrow = 2)
#LECTIN-A
#All populations
p1 <- filter(media_global_lectinvariation_df, Channels == "LECTIN-A", Subpopulation !=
"Dead + Apoptotic") %>%
mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S",
"Go/G1"))) %>%
mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
"MAL II"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
geom_point() +
geom_line(size = 1) +
theme_classic() +
labs(x = NULL , y = "LECTIN-A (linear scale)", title = NULL ) +
scale_colour_brewer(palette = "Dark2", name = "Subpopulation", guide = FALSE) +
scale_x_discrete(expand = c(0,0), breaks = c("a", "b", "c", "d", "e", "f", "g"), labels = c("-3", "-
2", "-1", "0", "+1", "+2", "+3")) +
facet_grid(.~ Lectin) +
theme_bw() +
theme(
   plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
```

```
1618      legend.text = element_text(size = 15),
1619      legend.title = element_text(size = 15, face = "bold"),
1620      legend.box.background = element_blank(),
1621      legend.justification = "center",
1622      legend.position = "bottom",
1623      #axis.text.x = element_text(angle = 45),
1624      #axis.ticks.x = element_blank(),
1625      axis.title = element_text(size = 15),
1626      strip.text = element_text(size = 15),
1627      strip.background = element_rect(fill = "grey90"),
1628      panel.grid = element_blank(),
1629      panel.spacing = unit(0.75, "lines")
1630      #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
1631      #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
1632      #legend.background = element_rect(fill = "grey90", colour = "grey90")
1633    )
1634  #LECTIN-A
1635  #DNA cycle populations
1636  p2 <- filter(media_global_lectinvariation_df, Channels == "LECTIN-A", Subpopulation !=
1637  "Dead + Apoptotic") %>%
1638  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S",
1639  "Go/G1"))) %>%
1640  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
1641  "MAL II"))) %>%
1642  ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
1643  geom_point() +
1644  geom_line(size = 1) +
1645  theme_classic() +
1646  labs(x = "Level of Spent medium (Days)" , y = "LECTIN-A (linear scale)", title = NULL ) +
1647  scale_colour_brewer(palette = "Dark2", name = "Subpopulation") +
```

```
scale_x_discrete(expand = c(0,0), breaks = c("a", "b", "c", "d", "e", "f", "g"), labels = c("-3", "-2", "-1", "0", "+1", "+2", "+3")) +
scale_y_continuous(limits = c(25, 180)) +
facet_grid(.~ Lectin) +
theme_bw() +
theme(
  plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
  legend.text = element_text(size = 15),
  legend.title = element_text(size = 15, face = "bold"),
  legend.box.background = element_blank(),
  legend.justification = "center",
  legend.position = "bottom",
  #axis.text.x = element_text(angle = 45),
  #axis.ticks.x = element_blank(),
  axis.title = element_text(size = 15),
  strip.text = element_text(size = 15),
  strip.background = element_rect(fill = "grey90"),
  panel.grid = element_blank(),
  panel.spacing = unit(0.75, "lines")
  #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
  #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
  #legend.background = element_rect(fill = "grey90", colour = "grey90")
 )
grid.arrange(p1, p2, nrow = 2)
```

Lectin Inferential Analysis

```{r}
Lectin_A_Subp_G2M_df <- table_manipulation(media_global_descriptive_df, media_global_F_T_df, c("LECTIN_A"), c("LECTIN-A"), c("G2/M"), c("Media"))
```

```
Lectin_A_Subp_S_df <- table_manipulation(media_global_descriptive_df,
media_global_F_T_df, c("LECTIN_A"), c("LECTIN-A"), c("S"), c("Media"))

Lectin_A_Subp_GoG1_df <- table_manipulation(media_global_descriptive_df,
media_global_F_T_df, c("LECTIN_A"), c("LECTIN-A"), c("Go/G1"), c("Media"))

Lectin_A_df <- rbind(Lectin_A_Subp_G2M_df, Lectin_A_Subp_S_df,
Lectin_A_Subp_GoG1_df)

Lectin_A_df$Lectin_face <- factor(Lectin_A_df$Lectin, levels = c("AAL", "LEC B", "PNA", "LEC
A", "AAL-2", "WGA", "MAL II"))

Lectin_A_df$Subpopulation_face <- factor(Lectin_A_df$Subpopulation, levels = c("G2/M",
"S", "Go/G1", "Apoptotic","Dead"))

media_global_descriptive_df$Lectin_face <- factor(media_global_descriptive_df$Lectin,
levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))

media_global_descriptive_df$Subpopulation_face <-
factor(media_global_descriptive_df$Subpopulation, levels = c("G2/M", "S", "Go/G1",
"Apoptotic","Dead"))

media_global_lectinvariation_df$Lectin_face <-
factor(media_global_lectinvariation_df$Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A",
"AAL-2", "WGA", "MAL II"))

#Lectin_A_df$Sample_Type_face <- factor(Lectin_A_df$Sample_Type, levels = c("6", "5",
"4", "3", "2", "1", "0"))

#media_global_descriptive_df$Sample_Type_face <-
factor(media_global_descriptive_df$Sample_Type, levels = c("6", "5", "4", "3", "2", "1",
"0"))

#media_global_lectinvariation_df$Sample_Type_face <-
factor(media_global_lectinvariation_df$Sample_Type, levels = c("6", "5", "4", "3", "2",
"1", "0"))

#set fill and colour manual

#d95f02 highly significant

#1b9e77 not significant

#7570b3 trend towards significance

#e7298a very highly significant

#66a61e significant

p_G2M  <- filter(Lectin_A_df, Subpopulation == "G2/M") %>%

ggplot(aes(Sample_Type, LECTIN_A)) +

  geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
```

```
1712     geom_boxplot(data = filter(media_global_descriptive_df, Sample_Type == '3',
1713   Subpopulation == "G2/M"), aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20,
1714   outlier.shape = NA) +
1715     facet_grid(Subpopulation ~ Lectin_face) +
1716     labs(x = "Level of Spent medium (Days)", y = "LECTIN-A (log scale)", title = NULL) +
1717       scale_x_discrete(breaks=c("0", "1", "2", "3", "4", "5", "6"),
1718                        labels=c("-3", "-2", "-1", "0", "+1", "+2", "+3")) +
1719       scale_fill_manual(name = "Level of Statistical Significance", values = c("#d95f02",
1720   "#1b9e77", "#66a61e", "#7570b3", "#e7298a")) +
1721       scale_colour_manual(values = c("#d95f02", "#1b9e77", "#66a61e", "#7570b3",
1722   "#e7298a"), guide = FALSE) +
1723       scale_y_continuous(expand = c(0,0)) +
1724    coord_cartesian(ylim = c(-80, 400)) +
1725   theme_bw() +
1726    theme(
1727      plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
1728      legend.text = element_text(size = 15),
1729      legend.title = element_text(size = 15, face = "bold"),
1730      legend.box.background = element_blank(),
1731      legend.justification = "center",
1732      legend.position = "bottom",
1733      #axis.text.x = element_text(angle = 45),
1734      #axis.ticks.x = element_blank(),
1735      axis.title = element_text(size = 15),
1736      strip.text = element_text(size = 15),
1737      strip.background = element_rect(fill = "grey90"),
1738      panel.grid = element_blank(),
1739      panel.spacing = unit(0.75, "lines")
1740      #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
1741      #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
1742      #legend.background = element_rect(fill = "grey90", colour = "grey90")
```

```
  )

df1 <- filter(media_global_descriptive_df, Sample_Type == 'd', Subpopulation %in%
c("Go/G1", "S", "G2/M"))

#d95f02 highly significant

#1b9e77 not significant

#7570b3 trend towards significance

#e7298a very highly significant

#66a61e significant

filter(Lectin_A_df, Subpopulation %in% c("Go/G1", "S", "G2/M")) %>%

ggplot(aes(Sample_Type, LECTIN_A)) +

  geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +

  geom_boxplot(data = filter(media_global_descriptive_df, Sample_Type == 'd',
Subpopulation %in% c("Go/G1", "S", "G2/M")), aes(Sample_Type, LECTIN_A), fill = "grey",
size = 0.20, outlier.shape = NA) +

  facet_grid(Subpopulation_face ~ Lectin_face) +

  labs(x = "Level of Spent medium (Days)", y = "LECTIN-A (linear scale)", title = NULL) +

  scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g"), labels = c("-3", "-2", "-1", "0",
"+1", "+2", "+3")) +

  scale_fill_manual(name = "Level of Statistical Significance", values = c("#d95f02",
"#1b9e77","#7570b3", "#e7298a")) +

   scale_colour_manual(values = c("#d95f02", "#1b9e77","#7570b3", "#e7298a"), guide =
FALSE) +

    scale_y_continuous(expand = c(0,0)) +

 coord_cartesian(ylim = c(-80, 400)) +

 theme_bw() +

  theme(

   plot.title = element_text(face = "bold", size = 18, hjust = 0.5),

   legend.text = element_text(size = 15),

   legend.title = element_text(size = 15, face = "bold"),

   legend.box.background = element_blank(),

   legend.justification = "center",
```

```
1774      legend.position = "bottom",
1775      #axis.text.x = element_text(angle = 45),
1776      #axis.ticks.x = element_blank(),
1777      axis.title = element_text(size = 15),
1778      strip.text = element_text(size = 15),
1779      strip.background = element_rect(fill = "grey90"),
1780      panel.grid = element_blank(),
1781      panel.spacing = unit(0.75, "lines")
1782      #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
1783      #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
1784      #legend.background = element_rect(fill = "grey90", colour = "grey90")
1785    )
1786  plot_line_box <-  filter(Lectin_A_df, Subpopulation == "Go/G1")  %>%
1787  ggplot(aes(Sample_Type, LECTIN_A)) +
1788    geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
1789    geom_boxplot(data = filter(media_global_descriptive_df, Sample_Type == 'd',
1790  Subpopulation == "Go/G1"), aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20,
1791  outlier.shape = NA) +
1792   #geom_point(data = filter(CO2_global_lectinvariation_df, Channels == "LECTIN-A",
1793  Subpopulation == "Go/G1"), aes(Sample_Type, Mean, group = Subpopulation), colour =
1794  "black", size = 0.6) +
1795   geom_line(data = filter(media_global_lectinvariation_df, Channels == "LECTIN-A",
1796  Subpopulation == "Go/G1"), aes(Sample_Type, Mean, group = Subpopulation), colour =
1797  "black", size = 0.6) +
1798    facet_grid(. ~ Lectin_face) +
1799    labs(x = "Level of Spent medium (Days)", y = "LECTIN-A (linear scale)", title = NULL) +
1800        scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g"), labels = c("-3", "-2", "-1", "0",
1801  "+1", "+2", "+3")) +
1802      scale_fill_manual(name = "Level of Statistical Significance", values = c("#1b9e77",
1803  "#e7298a")) +
1804      scale_colour_manual(values = c("#1b9e77", "#e7298a"), guide = FALSE) +
1805      scale_y_continuous(expand = c(0,0)) +
```

```
1806      coord_cartesian(ylim = c(-50, 270)) +
1807      theme_bw() +
1808       theme(
1809         plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
1810         legend.text = element_text(size = 15),
1811         legend.title = element_text(size = 15, face = "bold"),
1812         legend.box.background = element_blank(),
1813         legend.justification = "center",
1814         legend.position = "bottom",
1815         #axis.text.x = element_text(angle = 45),
1816         #axis.ticks.x = element_blank(),
1817         axis.title = element_text(size = 15),
1818         strip.text = element_text(size = 15),
1819         strip.background = element_rect(fill = "grey90"),
1820         panel.grid = element_blank(),
1821         panel.spacing = unit(0.75, "lines")
1822         #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
1823         #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
1824         #legend.background = element_rect(fill = "grey90", colour = "grey90")
1825       )
1826     df1 <- filter(media_global_descriptive_df, Sample_Type == '3', Subpopulation == "S")
1827     p_S <- filter(Lectin_A_df, Subpopulation == "S") %>%
1828     ggplot(aes(Sample_Type, LECTIN_A)) +
1829       geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
1830       geom_boxplot(data = df1, aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20,
1831     outlier.shape = NA) +
1832       facet_grid(Subpopulation ~ Lectin_face) +
1833       labs(x = "Level of Spent medium (Days)", y = "LECTIN-A (log scale)", title = NULL) +
1834        scale_x_discrete(breaks=c("0", "1", "2", "3", "4", "5", "6"),
```

```
                    labels=c("-3", "-2", "-1", "0", "+1", "+2", "+3")) +
      scale_fill_manual(name = "Level of Statistical Significance", values = c("#1b9e77",
"#e7298a")) +
      scale_colour_manual(values = c("#1b9e77", "#e7298a"), guide = FALSE) +
      scale_y_continuous(expand = c(0,0)) +
    coord_cartesian(ylim = c(-65, 350)) +
    theme_bw() +
     theme(
      plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
      legend.text = element_text(size = 15),
      legend.title = element_text(size = 15, face = "bold"),
      legend.box.background = element_blank(),
      legend.justification = "center",
      legend.position = "bottom",
      #axis.text.x = element_text(angle = 45),
      #axis.ticks.x = element_blank(),
      axis.title = element_text(size = 15),
      strip.text = element_text(size = 15),
      strip.background = element_rect(fill = "grey90"),
      panel.grid = element_blank(),
      panel.spacing = unit(0.75, "lines")
      #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
      #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
      #legend.background = element_rect(fill = "grey90", colour = "grey90")
     )
   df1 <- filter(media_global_descriptive_df, Sample_Type == '3', Subpopulation == "Go/G1")
   p_Go <- filter(Lectin_A_df, Subpopulation == "Go/G1") %>%
   ggplot(aes(Sample_Type, LECTIN_A)) +
     geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
```

```
1864     geom_boxplot(data = df1, aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20,
1865   outlier.shape = NA) +
1866    facet_grid(Subpopulation ~ Lectin_face) +
1867    labs(x = "Level of Spent medium (Days)", y = "LECTIN-A (log scale)", title = NULL) +
1868      scale_x_discrete(breaks=c("0", "1", "2", "3", "4", "5", "6"),
1869                       labels=c("-3", "-2", "-1", "0", "+1", "+2", "+3")) +
1870      scale_fill_manual(name = "Level of Statistical Significance", values = c("#1b9e77",
1871   "#e7298a")) +
1872      scale_colour_manual(values = c("#1b9e77", "#e7298a"), guide = FALSE) +
1873      scale_y_continuous(expand = c(0,0)) +
1874    coord_cartesian(ylim = c(-50, 280)) +
1875    theme_bw() +
1876     theme(
1877      plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
1878      legend.text = element_text(size = 15),
1879      legend.title = element_text(size = 15, face = "bold"),
1880      legend.box.background = element_blank(),
1881      legend.justification = "center",
1882      legend.position = "bottom",
1883      #axis.text.x = element_text(angle = 45),
1884      #axis.ticks.x = element_blank(),
1885      axis.title = element_text(size = 15),
1886      strip.text = element_text(size = 15),
1887      strip.background = element_rect(fill = "grey90"),
1888      panel.grid = element_blank(),
1889      panel.spacing = unit(0.75, "lines")
1890      #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
1891      #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
1892      #legend.background = element_rect(fill = "grey90", colour = "grey90")
1893     )
```

```

**Lectin Power Analysis**

```{r}

**#fd8d3c G2/M**

**#f03b20 S**

**#bd0026 Go/G1**

media_global_F_T_df$Lectin_face <- factor(media_global_F_T_df$Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))

filter(media_global_F_T_df, Channels == "LECTIN-A", Subpopulation %in% c("G2/M", "S", "Go/G1")) %>%

  mutate(Subpopulation = factor(Subpopulation, levels = c("G2/M", "S", "Go/G1"))) %>%

  mutate(Power = Power * 100) %>%

   ggplot(aes(Sample_Type, Power, fill = Subpopulation)) +

   geom_bar(stat = "identity", colour = NA) +

   facet_grid(Subpopulation ~ Lectin_face) +

   labs(x = "Level of Spent medium (Days)", y = "Power (%)", title = NULL) +

  **#scale_fill_brewer(palette = "RdBu", guide = FALSE) +**

    scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g"), labels = c("-3", "-2", "-1", "0", "+1", "+2", "+3")) +

   scale_fill_manual(values = c("#fd8d3c", "#f03b20", "#bd0026"), guide = FALSE) +

   scale_colour_manual(values = c("#fd8d3c", "#f03b20", "#bd0026"), guide = FALSE) +

   scale_y_continuous(expand = c(0,0), limits = c(0,100)) +

   theme_bw() +

  theme(

    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),

    legend.text = element_text(size = 15),

    legend.title = element_text(size = 15, face = "bold"),

    legend.box.background = element_blank(),

    legend.justification = "center",

    legend.position = "bottom",

```
      #axis.text.x = element_text(angle = 45),
      #axis.ticks.x = element_blank(),
      axis.title = element_text(size = 15),
      strip.text = element_text(size = 15),
      strip.background = element_rect(fill = "grey90"),
      panel.grid = element_blank(),
      panel.spacing = unit(0.75, "lines")
      #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
      #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
      #legend.background = element_rect(fill = "grey90", colour = "grey90")
)
  filter(media_global_F_T_df, Channels == "LECTIN-A", Subpopulation %in% c("G2/M", "S",
"Go/G1")) %>%
   mutate(Subpopulation = factor(Subpopulation, levels = c("G2/M", "S", "Go/G1"))) %>%
   ggplot(aes(Sample_Type, Sample_Size, fill = Subpopulation)) +
   geom_bar(stat = "identity", colour = NA) +
   facet_grid(Subpopulation ~ Lectin_face) +
   labs(x = "Level of Spent medium (Days)", y = "Sample size (number of cells)", title = NULL) +
   #scale_fill_brewer(palette = "RdBu", guide = FALSE) +
   scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g"), labels = c("-3", "-2", "-1", "0",
"+1", "+2", "+3")) +
   scale_fill_manual(values = c("#41b6c4", "#2c7fb8", "#253494"), guide = FALSE) +
   scale_colour_manual(values = c("#41b6c4", "#2c7fb8", "#253494"), guide = FALSE) +
   scale_y_continuous(expand = c(0,0)) +
   theme_bw() +
   theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
```

```
1954        legend.justification = "center",
1955        legend.position = "bottom",
1956        #axis.text.x = element_text(angle = 45),
1957        #axis.ticks.x = element_blank(),
1958        axis.title = element_text(size = 15),
1959        strip.text = element_text(size = 15),
1960        strip.background = element_rect(fill = "grey90"),
1961        panel.grid = element_blank(),
1962        panel.spacing = unit(0.75, "lines")
1963        #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
1964        #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
1965        #legend.background = element_rect(fill = "grey90", colour = "grey90")
1966      )
1967
1968    filter(media_global_F_T_df, Subpopulation =='S', Channels == "LECTIN-A") %>%
1969      mutate(Power = Power * 100) %>%
1970      ggplot(aes(Sample_Type, Power)) +
1971      geom_bar(stat = "identity" , colour = NA, fill = "#e7298a", alpha = 0.70) +
1972      facet_grid(Subpopulation ~ Lectin_face) +
1973      labs(x = "Level of Spent medium (Days)", y = "Power (%)", title = NULL) +
1974     #scale_fill_brewer(palette = "RdBu", guide = FALSE) +
1975     scale_x_discrete(breaks=c("0", "1", "2", "4", "5", "6"),
1976                      labels=c("-3", "-2", "-1", "+1", "+2", "+3")) +
1977     scale_y_continuous(expand = c(0,0), limits = c(0,100)) +
1978      theme_bw() +
1979     theme(
1980       plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
1981       legend.text = element_text(size = 15),
1982       legend.title = element_text(size = 15, face = "bold"),
```

```
      legend.box.background = element_blank(),
      legend.justification = "center",
      legend.position = "bottom",
      #axis.text.x = element_text(angle = 45),
      #axis.ticks.x = element_blank(),
      axis.title = element_text(size = 15),
      strip.text = element_text(size = 15),
      strip.background = element_rect(fill = "grey90"),
      panel.grid = element_blank(),
      panel.spacing = unit(0.75, "lines")
      #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
      #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
      #legend.background = element_rect(fill = "grey90", colour = "grey90")
    )
  geom_text(data = filter(media_global_F_T_df, Subpopulation =='Go/G1', Channels ==
"LECTIN-A") %>%
mutate(Power = Power * 100) %>% mutate_if(is.numeric, round, 1), aes(Sample_Type,
Power, label = Power), position = position_dodge(width = 0.8), size = 4, vjust = -0.5) +
dataMedian_PNA <- summarise(group_by(dfPNA, Sample), MD = median(Pacific.Blue.A))
%>% mutate_if(is.numeric, round, 1)
#fd8d3c G2/M
#f03b20 S
#bd0026 Go/G1
bar_plot <- filter(media_global_F_T_df, Subpopulation =='Go/G1', Channels == "LECTIN-A")
%>%
   mutate(Power = Power * 100) %>%
   ggplot(aes(Sample_Type, Power)) +
   geom_rect(aes(xmin = 0.4, xmax = 3.5, ymin = 0, ymax = Inf), fill = "#bd0026", alpha =
0.025) +
   geom_rect(aes(xmin = 3.5, xmax = Inf, ymin = 0, ymax = Inf), fill = "#bd0026", alpha = 0.07)
+
```

```
2014      geom_bar(stat = "identity" , colour = NA, fill = "#bd0026") +
2015      #geom_hline(aes(yintercept = c(25)), color = "grey70", linetype = "dashed", size = 1) +
2016      #geom_hline(aes(yintercept = c(50)), color = "grey50", linetype = "dashed", size = 1) +
2017      #geom_hline(aes(yintercept = c(75)), color = "grey30", linetype = "dashed", size = 1) +
2018      #geom_hline(aes(yintercept = c(90)), color = "black", linetype = "dotted", size = 1) +
2019     geom_text(data = filter(media_global_F_T_df, Subpopulation =='Go/G1', Channels ==
2020    "LECTIN-A") %>%
2021      mutate(Power = Power * 100) %>% mutate_if(is.numeric, round, 0), aes(Sample_Type,
2022    Power, label = Power), position = position_dodge(width = 0.8), size = 4, vjust = -0.5) +
2023      facet_grid(.~ Lectin_face) +
2024      labs(x = NULL, y = "Power (%)", title = NULL) +
2025     #scale_fill_brewer(palette = "RdBu", guide = FALSE) +
2026       scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g"), labels = c("-3", "-2", "-1", "0",
2027    "+1", "+2", "+3")) +
2028     scale_y_continuous(expand = c(0,0), limits = c(0,110), breaks = c(0, 25, 50, 75, 100)) +
2029      theme_bw() +
2030     theme(
2031       plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
2032       legend.text = element_text(size = 15),
2033       legend.title = element_text(size = 15, face = "bold"),
2034       legend.box.background = element_blank(),
2035       legend.justification = "center",
2036       legend.position = "bottom",
2037       #axis.text.x = element_text(angle = 45),
2038       #axis.ticks.x = element_blank(),
2039       axis.title = element_text(size = 15),
2040       strip.text = element_text(size = 15),
2041       strip.background = element_rect(fill = "grey90"),
2042       panel.grid = element_blank(),
2043       panel.spacing = unit(0.75, "lines")
```

```
    #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
    #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
    #legend.background = element_rect(fill = "grey90", colour = "grey90")
  )
grid.arrange(bar_plot, plot_line_box, nrow = 2)
filter(media_global_F_T_df, Channels == "LECTIN-A", Subpopulation %in% c("Go/G1", "S",
"G2/M")) %>%
  mutate(Power = Power * 100) %>%
  ggplot(aes(Sample_Type, Power, fill = Subpopulation)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_grid(. ~ Lectin_face) +
  labs(x = "Level of Spent medium (Days)", y = "Power (%)", title = NULL) +
  #scale_fill_brewer(palette = "RdBu", guide = FALSE) +
  scale_x_discrete(breaks=c("0", "1", "2", "3", "4", "5", "6"),
                   labels=c("-3", "-2", "-1","P", "+1", "+2", "+3")) +
  scale_y_continuous(expand = c(0,0), limits = c(0,100)) +
  theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    #axis.text.x = element_text(angle = 45),
    #axis.ticks.x = element_blank(),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
```

```
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
    #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
    #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
    #legend.background = element_rect(fill = "grey90", colour = "grey90")
  )
filter(media_global_ratios_df, Dimension %in% c('area', 'height', 'width')) %>%
  mutate(Level_of_sig_T = factor(Level_of_sig_T, levels = c("not significant", "trend toward significance", "significant", "highly significant", "very highly significant"))) %>%
  ggplot(aes(Sample_Type, Level_of_sig_T, fill = Comp_type)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_grid(Dimension ~ Lectin_face) +
  labs(x = "Media Depletion Levels (Days)", y = "Levels of Statistical significance", title = "Levels of Statistical Significance of SSC difference readings of G0/G1 cells across Media Depletion levels") +
  scale_fill_brewer(palette = "Dark2", name = "Levels of Statistical Significance") +

filter(media_global_F_T_df, Subpopulation =='Go/G1', Channels == "LECTIN-A") %>%
  mutate(Power = Power * 100) %>%
  group_by(Lectin) %>%
  summarise(Power_ave = mean(Power)) %>%
  mutate(Sample_Type = 3)
```

Analysis of Relative Lectin signal density

```{r}
library(gridExtra)
media_global_lectinvariation_df$Subpopulation_face <- factor(media_global_lectinvariation_df$Subpopulation, levels = c("Go/G1", "S", "G2/M", "Apoptotic","Dead"))
media_global_lectinvariation_df$Lectin_face <- factor(media_global_lectinvariation_df$Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))
```

```
p1 <- filter(media_global_lectinvariation_df, Channels %in% c("Area_ratio"),
Subpopulation_face %in% c("G2/M", "S", "Go/G1"), Lectin_face %in% c("AAL", "LEC B",
"PNA", "LEC A")) %>%
 ggplot(aes(Sample_Type, Mean, fill = Subpopulation_face, ymin = Mean - Mean_SD, ymax =
Mean + Mean_SD, group = Subpopulation_face)) +
 geom_bar(stat = "identity", position = "dodge") +
 geom_errorbar(size = 0.15, position = "dodge") +
  facet_grid(.~ Lectin_face) +
  labs(x = "Level of Spent medium (Day)", y = "LECTIN-A/FSC-A (linear scale)", title = NULL) +
  scale_fill_manual(values = c("#b2e2e2", "#66c2a4", "#238b45"), guide = FALSE) +
  scale_y_continuous(expand = c(0,0), limits = c(0,0.0025)) +
  scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g"), labels = c("-3", "-2", "-1", "0",
"+1", "+2", "+3")) +
  theme_bw() +
 theme(
   plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
   legend.text = element_text(size = 15),
   legend.title = element_text(size = 15, face = "bold"),
   legend.box.background = element_blank(),
   legend.justification = "center",
   legend.position = "bottom",
   axis.text.x = element_text(size = 10, colour = "black"),
   #axis.ticks.x = element_blank(),
   axis.title = element_text(size = 15),
   strip.text = element_text(size = 15),
   strip.background = element_rect(fill = "grey90"),
   panel.grid = element_blank()
   #panel.spacing = unit(0.75, "lines")
   #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
   #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
   #legend.background = element_rect(fill = "grey90", colour = "grey90")
```

```
2136     )
2137   p2 <- filter(media_global_lectinvariation_df, Channels %in% c("Area_ratio"),
2138   Subpopulation_face %in% c("G2/M", "S", "Go/G1"), Lectin_face %in% c("AAL-2", "WGA",
2139   "MAL II")) %>%
2140     ggplot(aes(Sample_Type, Mean, fill = Subpopulation_face, ymin = Mean - Mean_SD, ymax =
2141   Mean + Mean_SD, group = Subpopulation_face)) +
2142     geom_bar(stat = "identity", position = "dodge") +
2143     geom_errorbar(size = 0.15, position = "dodge") +
2144      facet_grid(.~ Lectin_face) +
2145      labs(x = "Level of Spent medium (Day)", y = "LECTIN-A/FSC-A (linear scale)", title = NULL) +
2146     scale_fill_manual(values = c("#b2e2e2", "#66c2a4", "#238b45"), name = "Subpopulation") +
2147      scale_y_continuous(expand = c(0,0), limits = c(0,0.0025)) +
2148      scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g"), labels = c("-3", "-2", "-1", "0",
2149   "+1", "+2", "+3")) +
2150      theme_bw() +
2151     theme(
2152       plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
2153       legend.text = element_text(size = 15),
2154       legend.title = element_text(size = 15, face = "bold"),
2155       legend.box.background = element_blank(),
2156       legend.justification = "center",
2157       legend.position = "right",
2158       axis.text.x = element_text(size = 10, colour = "black"),
2159       #axis.ticks.x = element_blank(),
2160       axis.title = element_text(size = 15),
2161       strip.text = element_text(size = 15),
2162       strip.background = element_rect(fill = "grey90"),
2163       panel.grid = element_blank()
2164       #panel.spacing = unit(0.75, "lines")
2165       #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
2166       #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
```

```
    #legend.background = element_rect(fill = "grey90", colour = "grey90")
  )
grid.arrange(p1, p2, nrow = 2)
```

8.3 Temperature data treatment and generation of plots

Data obtained from cells subjected to the variation of temperature levels are computed in this section. Pre built-in R functions and the functions created in Section 8.1 are used here. The code for the generation of plots are demonstrated in this section as well.

```
#Algorithm to collect and save gated data of Temperature Variation Experiments
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp & CO2 baseline")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp & CO2
baseline/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 32 C")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 32 C/Compensation
Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
```

```
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 33 C")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 33 C/Compensation
Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 34 C")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 34 C/Compensation
Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 35 C")
wd <- getwd()
```

```
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 35 C/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 36 C")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 36 C/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 38 C")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 38 C/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
```

```
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 39 C")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 39 C/Compensation
Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 40 C")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 40 C/Compensation
Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 41 C")
wd <- getwd()
```

```
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 41 C/Compensation Controls - WGA")

flow_gating_list <- flow_gating(wd, x_WGA)

#flow_gating_list <- flow_gating(wd)

gs1 <- flow_gating_list[[1]]

gs2 <- flow_gating_list[[2]]

gs3 <- flow_gating_list[[3]]

save_gs(gs1, path = file.path(wd, "gs1"))

save_gs(gs2, path = file.path(wd, "gs2"))

save_gs(gs3, path = file.path(wd, "gs3"))
```

#Algorithm to retrieve gated data and run the statistical analysis of the Temperature Variation Experiments

```{r}
#Algorithm to process Temperature Variation Experiments

setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp & CO2 baseline")

wd <- getwd()

gs1 <- load_gs(file.path(wd, "gs1"))

gs2 <- load_gs(file.path(wd, "gs2"))

gs3 <- load_gs(file.path(wd, "gs3"))

temp_baseline <- table_summary(gs1, gs2, gs3, c("37"))

temp_baseline_descriptive <- table_descriptive(gs1, gs2, gs3, c("37"))

temp_baseline_density <- lectin_density_stats(temp_baseline)

setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 32 C")

wd <- getwd()

gs1 <- load_gs(file.path(wd, "gs1"))

gs2 <- load_gs(file.path(wd, "gs2"))

gs3 <- load_gs(file.path(wd, "gs3"))

temp_32 <- table_summary(gs1, gs2, gs3, c("32"))

temp_32_descriptive <- table_descriptive(gs1, gs2, gs3, c("32"))
```

```
base_32 <- F_T_Power_test(temp_baseline, temp_32, c("32"))
temp_32_density <- lectin_density_stats(temp_32)
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 33 C")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
temp_33 <- table_summary(gs1, gs2, gs3, c("33"))
temp_33_descriptive <- table_descriptive(gs1, gs2, gs3, c("33"))
base_33 <- F_T_Power_test(temp_baseline, temp_33, c("33"))
temp_33_density <- lectin_density_stats(temp_33)
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 34 C")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
temp_34 <- table_summary(gs1, gs2, gs3, c("34"))
temp_34_descriptive <- table_descriptive(gs1, gs2, gs3, c("34"))
base_34 <- F_T_Power_test(temp_baseline, temp_34, c("34"))
temp_34_density <- lectin_density_stats(temp_34)
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 35 C")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
temp_35 <- table_summary(gs1, gs2, gs3, c("35"))
temp_35_descriptive <- table_descriptive(gs1, gs2, gs3, c("35"))
base_35 <- F_T_Power_test(temp_baseline, temp_35, c("35"))
temp_35_density <- lectin_density_stats(temp_35)
```

```
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 36 C")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
temp_36 <- table_summary(gs1, gs2, gs3, c("36"))
temp_36_descriptive <- table_descriptive(gs1, gs2, gs3, c("36"))
base_36 <- F_T_Power_test(temp_baseline, temp_36, c("36"))
temp_36_density <- lectin_density_stats(temp_36)
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 38 C")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
temp_38 <- table_summary(gs1, gs2, gs3, c("38"))
temp_38_descriptive <- table_descriptive(gs1, gs2, gs3, c("38"))
base_38 <- F_T_Power_test(temp_baseline, temp_38, c("38"))
temp_38_density <- lectin_density_stats(temp_38)
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 39 C")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
temp_39 <- table_summary(gs1, gs2, gs3, c("39"))
temp_39_descriptive <- table_descriptive(gs1, gs2, gs3, c("39"))
base_39 <- F_T_Power_test(temp_baseline, temp_39, c("39"))
temp_39_density <- lectin_density_stats(temp_39)
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation/Temp 40 C")
wd <- getwd()
```

```
gs1 <- load_gs(file.path(wd, "gs1"))

gs2 <- load_gs(file.path(wd, "gs2"))

gs3 <- load_gs(file.path(wd, "gs3"))

temp_40 <- table_summary(gs1, gs2, gs3, c("40"))

temp_40_descriptive <- table_descriptive(gs1, gs2, gs3, c("40"))

base_40 <- F_T_Power_test(temp_baseline, temp_40, c("40"))

temp_40_density <- lectin_density_stats(temp_40)

temp_global_descriptive_df <- rbind(temp_baseline_descriptive, temp_32_descriptive,
temp_33_descriptive, temp_34_descriptive, temp_35_descriptive, temp_36_descriptive,
temp_38_descriptive, temp_39_descriptive, temp_40_descriptive)

rm(temp_baseline_descriptive, temp_32_descriptive, temp_33_descriptive,
temp_34_descriptive, temp_35_descriptive, temp_36_descriptive, temp_38_descriptive,
temp_39_descriptive, temp_40_descriptive)

temp_global_lectinvariation_df <- rbind(temp_baseline, temp_32, temp_33, temp_34,
temp_35, temp_36, temp_38, temp_39, temp_40)

rm(temp_baseline, temp_32, temp_33, temp_34, temp_35, temp_36, temp_38, temp_39,
temp_40)

temp_global_F_T_df <- rbind(base_32, base_33, base_34, base_35, base_36, base_38,
base_39, base_40)

rm(base_32, base_33, base_34, base_35, base_36, base_38, base_39, base_40)

temp_global_density_df <- rbind(temp_baseline_density, temp_32_density,
temp_33_density, temp_34_density, temp_35_density, temp_36_density,
temp_38_density, temp_39_density, temp_40_density)

rm(temp_baseline_density, temp_32_density, temp_33_density, temp_34_density,
temp_35_density, temp_36_density, temp_38_density, temp_39_density,
temp_40_density)

colnames(temp_global_F_T_df) <- c("Sample_Type", "Channels", "Subpopulation", "Lectin",
"Mean", "SD", "Sample_Size","Fp_value", "F_significance", "F_test_conclusion", "Tp_value",
"T_test_significance", "Power")

colnames(temp_global_lectinvariation_df) <- c("Sample_Type", "Channels", "Mean",
"Mean_SD", "CV_perc", "Subpopulation", "Sample_Size","Viability_perc",
"Viability_SD_perc", "Lectin")
```

```{r}
library(readxl)
```

```
library(gridExtra)
#Read in the excel spreadsheet into R
setwd("~/Dropbox/PhD Project/PhD Project/Temp Variation")
pH_temp <- read_excel("pH.xlsx")
pH_temp_df <- as.data.frame(pH_temp, stringsAsFactors = FALSE)
```

Viability and pH Plots

```{r}
#Viability across temperature variation (line plot of individual lectin curves)
viability_plot <- ggplot(temp_global_lectinvariation_df, aes(Sample_Type, Viability_perc)) +
  geom_smooth(aes(group = Lectin, color = Lectin), size = 1.5, se = FALSE) +
  scale_colour_manual(name = "Lectin",  values = c("#980043","#7a0177", "#08519c",
"#006d2c", "#7fcdbb",  "#ff7f00", "#993404")) +
   labs(x = "Temperature (ºC)" , y = "Viability (%)", title = NULL ) +
scale_x_discrete(expand = c(0,0), breaks = c("32", "33", "34", "35", "36", "37", "38", "39",
"40"), labels = c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
  theme_classic() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "right",
    #axis.text.x = element_text(size = 10, face = "bold", color = "black"),
    #axis.ticks.x = element_blank(),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90")
   )
```

```
#pH line plot across temperature variation
pH_plot <- ggplot(pH_temp_df, aes(Sample_Type, pH)) +
  geom_point(aes(colour = Replicate), size = 3, alpha = 0.60) +
  geom_smooth(size = 1.5) +
  labs(x = "Temperature (ºC)" , y = "pH", title = NULL ) +
  scale_colour_brewer(palette = "Set1", name = "Replicate") +
  scale_x_continuous(expand = c(0,0), breaks = c(32, 33, 34, 35, 36, 37, 38, 39, 40), labels =
c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
  scale_y_continuous(expand = c(0,0), limits = c(6.8, 8)) +
  theme_classic() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "right",
    #axis.text.x = element_text(size = 10, face = "bold", color = "black"),
    #axis.ticks.x = element_blank(),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90")
  )
ggplot(media_global_lectinvariation_df, aes(Sample_Type, Viability_perc)) +
  geom_point(alpha = 0.0) +
 geom_smooth(data = viabilityPNA_df, aes(colour = "A"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 8), se = FALSE) +
  geom_smooth(data = viabilityAAL_df, aes(colour = "B"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 7), se = FALSE) +
  geom_smooth(data = viabilityMALII_df, aes(colour = "C"), method = "lm", size = 1.5,
formula = y ~ splines::bs(x, 8), se = FALSE) +
```

```
  geom_smooth(data = viabilityLECB_df, aes(colour = "D"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 8), se = FALSE) +

  geom_smooth(data = viabilityLECA_df, aes(colour = "E"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 8), se = FALSE) +

  geom_smooth(data = viabilityAAL2_df, aes(colour = "F"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 8), se = FALSE) +

  geom_smooth(data = viabilityWGA_df, aes(colour = "G"), method = "lm", size = 1.5,
formula = y ~ splines::bs(x, 6), se = FALSE) +

geom_vline(aes(xintercept = c(6.95)), color = "red", linetype = "dashed", size = 1) +

#geom_text(aes(x = 5.5, label = expression(paste("Lectin concentration level\n selected at
3.0", mu, "g/mL")) , y = 96), colour="red", angle = 0) +

  labs(x = expression(paste("Lectin concentration (", mu, "g/mL")) , y = "Viability (%)", title =
NULL ) +

  scale_colour_manual(name = "Lectin",  values = c("#ff7f00", "#980043", "#7fcdbb",
"#006d2c", "#08519c", "#7a0177", "#993404"), breaks = c("A", "B", "C", "D", "E", "F", "G"),
labels = c("PNA", "AAL", "MALII", "LECB", "LECA", "AAL-2", "WGA")) +

scale_x_discrete(expand = c(0,0)) +

  theme_classic() +

  theme(

    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),

    legend.text = element_text(size = 15),

    legend.title = element_text(size = 15, face = "bold"),

    legend.box.background = element_blank(),

    legend.justification = "center",

    legend.position = "right",

    #axis.text.x = element_text(size = 10, face = "bold", color = "black"),

    #axis.ticks.x = element_blank(),

    axis.title = element_text(size = 15),

    strip.text = element_text(size = 15),

    strip.background = element_rect(fill = "grey90"),

    #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),

    #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
```

```
    #legend.background = element_rect(fill = "grey90", colour = "grey90")
  )
```
```
Facetted plots with all lectins - Descriptive Analysis
```{r}
#FSC-A
#AAL, LECB, PNA, LECA, AAL-2, WGA, MAL II
filter(temp_global_lectinvariation_df, Channels == "FSC-A", Subpopulation != "Dead + Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S", "Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
  geom_point() +
  geom_line(size = 1) +
  theme_classic() +
  labs(x = "Temperature (ºC)"  , y = "FSC-A (linear scale)", title = NULL ) +
  scale_colour_brewer(palette = "Dark2", name = "Subpopulation") +
scale_x_discrete(expand = c(0,0), breaks = c("32", "33", "34", "35", "36", "37", "38", "39", "40"), labels = c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
  facet_grid(.~ Lectin) +
  theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
```
```

```
    #axis.text.x = element_text(angle = 45),
    #axis.ticks.x = element_blank(),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
    #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
    #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
    #legend.background = element_rect(fill = "grey90", colour = "grey90")
  )
#SSC-A
#All populations
p1 <- filter(temp_global_lectinvariation_df, Channels == "SSC-A", Subpopulation != "Dead +
Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S",
"Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
"MAL II"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
  geom_point() +
  geom_line(size = 1) +
  theme_classic() +
  labs(x = NULL , y = "SSC-A (linear scale)", title = NULL ) +
  scale_colour_brewer(palette = "Dark2", name = "Subpopulation", guide = FALSE) +
scale_x_discrete(expand = c(0,0), breaks = c("32", "33", "34", "35", "36", "37", "38", "39",
"40"), labels = c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
  facet_grid(.~ Lectin) +
  theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
```

```
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    #axis.text.x = element_text(angle = 45),
    #axis.ticks.x = element_blank(),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
    #panel.grid.major = element_line(size = 0.25, linetype = 'solid', colour = "grey90"),
    #panel.grid.minor = element_line(size = 0.125, linetype = 'solid', colour = "grey90")
    #legend.background = element_rect(fill = "grey90", colour = "grey90")
  )
#SSC-A
#DNA cycle populations
p2 <- filter(temp_global_lectinvariation_df, Channels == "SSC-A", Subpopulation != "Dead + Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S", "Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
  geom_point() +
  geom_line(size = 1) +
  theme_classic() +
  labs(x = "Temperature (ºC)" , y = "SSC-A (linear scale)", title = NULL ) +
  scale_colour_brewer(palette = "Dark2", name = "Subpopulation") +
```

```
scale_x_discrete(expand = c(0,0), breaks = c("32", "33", "34", "35", "36", "37", "38", "39",
"40"), labels = c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
  scale_y_continuous(limits = c(20000, 55000)) +
  facet_grid(.~ Lectin) +
  theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
   )
grid.arrange(p1, p2, nrow = 2)
#LECTIN-A
#All populations
p1 <- filter(temp_global_lectinvariation_df, Channels == "LECTIN-A", Subpopulation != "Dead
+ Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S",
"Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
"MAL II"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
  geom_point() +
  geom_line(size = 1) +
  theme_classic() +
```

```
  labs(x = NULL , y = "LECTIN-A (linear scale)", title = NULL ) +
  scale_colour_brewer(palette = "Dark2", name = "Subpopulation", guide = FALSE) +
 scale_x_discrete(expand = c(0,0), breaks = c("32", "33", "34", "35", "36", "37", "38", "39",
"40"), labels = c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
  facet_grid(.~ Lectin) +
  theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
   )
#LECTIN-A
#DNA cycle populations
p2 <- filter(temp_global_lectinvariation_df, Channels == "LECTIN-A", Subpopulation != "Dead
+ Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S",
"Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
"MAL II"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
  geom_point() +
  geom_line(size = 1) +
  theme_classic() +
```

```
    labs(x = "Temperature (ºC)", y = "LECTIN-A (linear scale)", title = NULL ) +
    scale_colour_brewer(palette = "Dark2", name = "Subpopulation") +
    scale_x_discrete(expand = c(0,0), breaks = c("32", "33", "34", "35", "36", "37", "38", "39",
"40"), labels = c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
    scale_y_continuous(limits = c(40, 240)) +
    facet_grid(.~ Lectin) +
    theme_bw() +
    theme(
      plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
      legend.text = element_text(size = 15),
      legend.title = element_text(size = 15, face = "bold"),
      legend.box.background = element_blank(),
      legend.justification = "center",
      legend.position = "bottom",
      axis.title = element_text(size = 15),
      strip.text = element_text(size = 15),
      strip.background = element_rect(fill = "grey90"),
      panel.grid = element_blank(),
      panel.spacing = unit(0.75, "lines")
     )
  grid.arrange(p1, p2, nrow = 2)
  ```
  Lectin Inferential Analysis
  ```{r}
  Lectin_A_Subp_G2M_df <- table_manipulation(temp_global_descriptive_df,
  temp_global_F_T_df, c("LECTIN_A"), c("LECTIN-A"), c("G2/M"), c("Temp"))
  Lectin_A_Subp_S_df <- table_manipulation(temp_global_descriptive_df,
  temp_global_F_T_df, c("LECTIN_A"), c("LECTIN-A"), c("S"), c("Temp"))
  Lectin_A_Subp_GoG1_df <- table_manipulation(temp_global_descriptive_df,
  temp_global_F_T_df, c("LECTIN_A"), c("LECTIN-A"), c("Go/G1"), c("Temp"))
```

```
Lectin_A_df <- rbind(Lectin_A_Subp_G2M_df, Lectin_A_Subp_S_df,
Lectin_A_Subp_GoG1_df)

Lectin_A_df$Lectin_face <- factor(Lectin_A_df$Lectin, levels = c("AAL", "LEC B", "PNA", "LEC
A", "AAL-2", "WGA", "MAL II"))

Lectin_A_df$Subpopulation_face <- factor(Lectin_A_df$Subpopulation, levels = c("G2/M",
"S", "Go/G1", "Apoptotic","Dead"))

temp_global_descriptive_df$Lectin_face <- factor(temp_global_descriptive_df$Lectin, levels
= c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))

temp_global_descriptive_df$Subpopulation_face <-
factor(temp_global_descriptive_df$Subpopulation, levels = c("G2/M", "S", "Go/G1",
"Apoptotic","Dead"))

temp_global_lectinvariation_df$Lectin_face <-
factor(temp_global_lectinvariation_df$Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A",
"AAL-2", "WGA", "MAL II"))

#set fill and colour manual

#d95f02 highly significant

#1b9e77 not significant

#7570b3 trend towards significance

#e7298a very highly significant

#66a61e significant

p_G2M  <- filter(Lectin_A_df, Subpopulation == "G2/M") %>%

ggplot(aes(Sample_Type, LECTIN_A)) +

  geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +

  geom_boxplot(data = filter(temp_global_descriptive_df, Sample_Type == '37',
Subpopulation == "G2/M"), aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20,
outlier.shape = NA) +

  facet_grid(.~ Lectin_face) +

  labs(x = "Temperature (ºC)", y = "LECTIN-A (log scale)", title = NULL) +

   scale_x_discrete(expand = c(0,0), breaks = c("32", "33", "34", "35", "36", "37", "38", "39",
"40"), labels = c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +

    scale_fill_manual(name = "Level of Statistical Significance", values = c("#1b9e77",
"#e7298a")) +

    scale_colour_manual(values = c("#1b9e77", "#e7298a"), guide = FALSE) +

    scale_y_continuous(expand = c(0,0)) +
```

```
2715      coord_cartesian(ylim = c(-50, 480)) +
2716     theme_bw() +
2717      theme(
2718        plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
2719        legend.text = element_text(size = 15),
2720        legend.title = element_text(size = 15, face = "bold"),
2721        legend.box.background = element_blank(),
2722        legend.justification = "center",
2723        legend.position = "bottom",
2724        axis.title = element_text(size = 15),
2725        strip.text = element_text(size = 15),
2726        strip.background = element_rect(fill = "grey90"),
2727        panel.grid = element_blank(),
2728        panel.spacing = unit(0.75, "lines")
2729       )
2730
2731   df1 <- filter(media_global_descriptive_df, Sample_Type == '37', Subpopulation %in%
2732   c("Go/G1", "S", "G2/M"))
2733   filter(Lectin_A_df, Subpopulation %in% c("Go/G1", "S", "G2/M")) %>%
2734   ggplot(aes(Sample_Type, LECTIN_A)) +
2735     geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
2736     geom_boxplot(data = filter(temp_global_descriptive_df, Sample_Type == '37',
2737   Subpopulation %in% c("Go/G1", "S", "G2/M")), aes(Sample_Type, LECTIN_A), fill = "grey",
2738   size = 0.20, outlier.shape = NA) +
2739     facet_grid(Subpopulation_face ~ Lectin_face) +
2740     labs(x = "Temperature (ºC)", y = "LECTIN-A (linear scale)", title = NULL) +
2741      scale_x_discrete(breaks = c("32", "33", "34", "35", "36", "37", "38", "39", "40"), labels =
2742   c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
2743       scale_fill_manual(name = "Level of Statistical Significance", values = c("#1b9e77",
2744   "#e7298a")) +
2745       scale_colour_manual(values = c("#1b9e77", "#e7298a"), guide = FALSE) +
```

```
2746        scale_y_continuous(expand = c(0,0)) +
2747      coord_cartesian(ylim = c(-80, 480)) +
2748      theme_bw() +
2749       theme(
2750         plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
2751         legend.text = element_text(size = 15),
2752         legend.title = element_text(size = 15, face = "bold"),
2753         legend.box.background = element_blank(),
2754         legend.justification = "center",
2755         legend.position = "bottom",
2756         axis.title = element_text(size = 15),
2757         strip.text = element_text(size = 15),
2758         strip.background = element_rect(fill = "grey90"),
2759         panel.grid = element_blank(),
2760         panel.spacing = unit(0.75, "lines")
2761        )
2762    plot_line_box <-  filter(Lectin_A_df, Subpopulation == "Go/G1")  %>%
2763    ggplot(aes(Sample_Type, LECTIN_A)) +
2764    geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
2765    geom_boxplot(data = filter(temp_global_descriptive_df, Sample_Type == '37',
2766    Subpopulation == "Go/G1"), aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20,
2767    outlier.shape = NA) +
2768      geom_line(data = filter(temp_global_lectinvariation_df, Channels == "LECTIN-A",
2769    Subpopulation == "Go/G1"), aes(Sample_Type, Mean, group = Subpopulation), colour =
2770    "black", size = 0.6) +
2771       facet_grid(.~ Lectin_face) +
2772       labs(x = "Temperature (ºC)", y = "LECTIN-A (linear scale)", title = NULL) +
2773        scale_x_discrete(breaks = c("32", "33", "34", "35", "36", "37", "38", "39", "40"), labels =
2774    c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
2775         scale_fill_manual(name = "Level of Statistical Significance", values = c("#1b9e77",
2776    "#e7298a")) +
```

```
2777        scale_colour_manual(values = c("#1b9e77", "#e7298a"), guide = FALSE) +
2778        scale_y_continuous(expand = c(0,0)) +
2779    coord_cartesian(ylim = c(-30, 350)) +
2780     theme_bw() +
2781      theme(
2782        plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
2783        legend.text = element_text(size = 15),
2784        legend.title = element_text(size = 15, face = "bold"),
2785        legend.box.background = element_blank(),
2786        legend.justification = "center",
2787        legend.position = "bottom",
2788        axis.title = element_text(size = 15),
2789        strip.text = element_text(size = 15),
2790        strip.background = element_rect(fill = "grey90"),
2791        panel.grid = element_blank(),
2792        panel.spacing = unit(0.75, "lines")
2793       )
2794    df1 <- filter(media_global_descriptive_df, Sample_Type == '3', Subpopulation == "S")
2795    p_S <- filter(Lectin_A_df, Subpopulation == "S") %>%
2796    ggplot(aes(Sample_Type, LECTIN_A)) +
2797      geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
2798      geom_boxplot(data = df1, aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20,
2799    outlier.shape = NA) +
2800      facet_grid(Subpopulation ~ Lectin_face) +
2801      labs(x = "Level of Spent medium (Days)", y = "LECTIN-A (log scale)", title = NULL) +
2802        scale_x_discrete(breaks=c("0", "1", "2", "3", "4", "5", "6"),
2803                         labels=c("-3", "-2", "-1", "0", "+1", "+2", "+3")) +
2804        scale_fill_manual(name = "Level of Statistical Significance", values = c("#1b9e77",
2805    "#e7298a")) +
2806        scale_colour_manual(values = c("#1b9e77", "#e7298a"), guide = FALSE) +
```

```
      scale_y_continuous(expand = c(0,0)) +
    coord_cartesian(ylim = c(-65, 350)) +
    theme_bw() +
     theme(
       plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
       legend.text = element_text(size = 15),
       legend.title = element_text(size = 15, face = "bold"),
       legend.box.background = element_blank(),
       legend.justification = "center",
       legend.position = "bottom",
       axis.title = element_text(size = 15),
       strip.text = element_text(size = 15),
       strip.background = element_rect(fill = "grey90"),
       panel.grid = element_blank(),
       panel.spacing = unit(0.75, "lines")
      )
```

#d95f02 highly significant

#1b9e77 not significant

#7570b3 trend towards significance

#e7298a very highly significant

#66a61e significant

```
```

Lectin Power Analysis

```{r}
```

#fd8d3c G2/M

#f03b20 S

#bd0026 Go/G1

```
temp_global_F_T_df$Lectin_face <- factor(temp_global_F_T_df$Lectin, levels = c("AAL",
"LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))
```

```
bar_plot <- filter(temp_global_F_T_df, Subpopulation =='Go/G1', Channels == "LECTIN-A") %>%
  mutate(Power = Power * 100) %>%
  ggplot(aes(Sample_Type, Power)) +
  geom_rect(aes(xmin = 0.4, xmax = 5.5, ymin = 0, ymax = Inf), fill = "#bd0026", alpha = 0.025) +
  geom_rect(aes(xmin = 5.5, xmax = Inf, ymin = 0, ymax = Inf), fill = "#bd0026", alpha = 0.07) +
  geom_bar(stat = "identity" , colour = NA, fill = "#bd0026") +
  #geom_hline(aes(yintercept = c(25)), color = "grey70", linetype = "dashed", size = 1) +
  #geom_hline(aes(yintercept = c(50)), color = "grey50", linetype = "dashed", size = 1) +
  #geom_hline(aes(yintercept = c(75)), color = "grey30", linetype = "dashed", size = 1) +
  #geom_hline(aes(yintercept = c(90)), color = "black", linetype = "dotted", size = 1) +
 geom_text(data = filter(temp_global_F_T_df, Subpopulation =='Go/G1', Channels == "LECTIN-A") %>%
  mutate(Power = Power * 100) %>% mutate_if(is.numeric, round, 0), aes(Sample_Type, Power, label = Power), position = position_dodge(width = 0.8), size = 4, vjust = -0.5) +
  facet_grid(.~ Lectin_face) +
  labs(x = NULL, y = "Power (%)", title = NULL) +
   scale_x_discrete(breaks = c("32", "33", "34", "35", "36", "37", "38", "39", "40"), labels = c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
 scale_y_continuous(expand = c(0,0), limits = c(0,110), breaks = c(0, 25, 50, 75, 100)) +
  theme_bw() +
 theme(
  plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
  legend.text = element_text(size = 15),
  legend.title = element_text(size = 15, face = "bold"),
  legend.box.background = element_blank(),
  legend.justification = "center",
  legend.position = "bottom",
  axis.title = element_text(size = 15),
```

```
      strip.text = element_text(size = 15),
      strip.background = element_rect(fill = "grey90"),
      panel.grid = element_blank(),
      panel.spacing = unit(0.75, "lines")
     )
grid.arrange(bar_plot, plot_line_box, nrow = 2)
filter(temp_global_F_T_df, Channels == "LECTIN-A", Subpopulation %in% c("G2/M", "S",
"Go/G1")) %>%
 mutate(Subpopulation = factor(Subpopulation, levels = c("G2/M", "S", "Go/G1"))) %>%
 mutate(Power = Power * 100) %>%
 ggplot(aes(Sample_Type, Power, fill = Subpopulation)) +
 geom_bar(stat = "identity", colour = NA) +
 facet_grid(Subpopulation ~ Lectin_face) +
 labs(x = "Temperature (ºC)", y = "Power (%)", title = NULL) +
  scale_x_discrete(expand = c(0,0), breaks = c("32", "33", "34", "35", "36", "37", "38", "39",
"40"), labels = c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
 scale_fill_manual(values = c("#fd8d3c", "#f03b20", "#bd0026"), guide = FALSE) +
 scale_colour_manual(values = c("#fd8d3c", "#f03b20", "#bd0026"), guide = FALSE) +
 scale_y_continuous(expand = c(0,0), limits = c(0,100)) +
 theme_bw() +
 theme(
  plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
  legend.text = element_text(size = 15),
  legend.title = element_text(size = 15, face = "bold"),
  legend.box.background = element_blank(),
  legend.justification = "center",
  legend.position = "bottom",
  axis.title = element_text(size = 15),
  strip.text = element_text(size = 15),
  strip.background = element_rect(fill = "grey90"),
```

```
2897        panel.grid = element_blank(),
2898        panel.spacing = unit(0.75, "lines")
2899       )
2900     filter(temp_global_F_T_df, Channels == "LECTIN-A", Subpopulation %in% c("G2/M", "S",
2901     "Go/G1")) %>%
2902      mutate(Subpopulation = factor(Subpopulation, levels = c("G2/M", "S", "Go/G1"))) %>%
2903       ggplot(aes(Sample_Type, Sample_Size, fill = Subpopulation)) +
2904       geom_bar(stat = "identity", colour = NA) +
2905       facet_grid(Subpopulation ~ Lectin_face) +
2906       labs(x = "Temperature (ºC)", y = "Sample size (number of cells)", title = NULL) +
2907      #scale_fill_brewer(palette = "RdBu", guide = FALSE) +
2908      scale_x_discrete(expand = c(0,0), breaks = c("32", "33", "34", "35", "36", "37", "38", "39",
2909     "40"), labels = c("-5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
2910       scale_fill_manual(values = c("#41b6c4", "#2c7fb8", "#253494"), guide = FALSE) +
2911       scale_colour_manual(values = c("#41b6c4", "#2c7fb8", "#253494"), guide = FALSE) +
2912       scale_y_continuous(expand = c(0,0)) +
2913       theme_bw() +
2914      theme(
2915        plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
2916        legend.text = element_text(size = 15),
2917        legend.title = element_text(size = 15, face = "bold"),
2918        legend.box.background = element_blank(),
2919        legend.justification = "center",
2920        legend.position = "bottom",
2921        axis.title = element_text(size = 15),
2922        strip.text = element_text(size = 15),
2923        strip.background = element_rect(fill = "grey90"),
2924        panel.grid = element_blank(),
2925        panel.spacing = unit(0.75, "lines")
2926       )
```

```

Analysis of Relative Lectin signal density

```{r}

library(gridExtra)

temp_global_lectinvariation_df$Subpopulation_face <-
factor(temp_global_lectinvariation_df$Subpopulation, levels = c("Go/G1", "S", "G2/M",
"Apoptotic","Dead"))

temp_global_lectinvariation_df$Lectin_face <-
factor(temp_global_lectinvariation_df$Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A",
"AAL-2", "WGA", "MAL II"))

**#b2e2e2 Go/G1**

**#66c2a4 S**

**#238b45 G2/M**

p1 <- filter(temp_global_lectinvariation_df, Channels %in% c("Area_ratio"),
Subpopulation_face %in% c("G2/M", "S", "Go/G1"), Lectin_face %in% c("AAL", "LEC B",
"PNA", "LEC A")) %>%

 ggplot(aes(Sample_Type, Mean, fill = Subpopulation_face, ymin = Mean - Mean_SD, ymax =
Mean + Mean_SD, group = Subpopulation_face)) +

 geom_bar(stat = "identity", position = "dodge") +

 geom_errorbar(size = 0.15, position = "dodge") +

  facet_grid(.~ Lectin_face) +

  labs(x = "Temperature (ºC)", y = "LECTIN-A/FSC-A (linear scale)", title = NULL) +

  scale_fill_manual(values = c("#b2e2e2", "#66c2a4", "#238b45"), guide = FALSE) +

  scale_y_continuous(expand = c(0,0), limits = c(0,0.0025)) +

 scale_x_discrete(breaks = c("32", "33", "34", "35", "36", "37", "38", "39", "40"), labels = c("-
5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +

  theme_bw() +

 theme(

   plot.title = element_text(face = "bold", size = 18, hjust = 0.5),

   legend.text = element_text(size = 15),

   legend.title = element_text(size = 15, face = "bold"),

   legend.box.background = element_blank(),

```
    legend.justification = "center",
    legend.position = "bottom",
    axis.text.x = element_text(size = 10, colour = "black"),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank()

)
p2 <- filter(temp_global_lectinvariation_df, Channels %in% c("Area_ratio"),
Subpopulation_face %in% c("G2/M", "S", "Go/G1"), Lectin_face %in% c("AAL-2", "WGA",
"MAL II")) %>%
  ggplot(aes(Sample_Type, Mean, fill = Subpopulation_face, ymin = Mean - Mean_SD, ymax =
Mean + Mean_SD, group = Subpopulation_face)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(size = 0.15, position = "dodge") +
   facet_grid(.~ Lectin_face) +
   labs(x = "Temperature (ºC)", y = "LECTIN-A/FSC-A (linear scale)", title = NULL) +
  scale_fill_manual(values = c("#b2e2e2", "#66c2a4", "#238b45"), name = "Subpopulation") +
   scale_y_continuous(expand = c(0,0), limits = c(0,0.0030)) +
 scale_x_discrete(breaks = c("32", "33", "34", "35", "36", "37", "38", "39", "40"), labels = c("-
5", "-4", "-3", "-2", "-1", "0", "+1", "+2", "+3")) +
   theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "right",
    axis.text.x = element_text(size = 10, colour = "black"),
```

```
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank()
  )
grid.arrange(p1, p2, nrow = 2)
```

8.4 CO_2 data treatment and generation of plots

Data obtained from cells subjected to the variation of CO_2 levels are computed in this section. Pre built-in R functions and the functions created in Section 8.1 are used here. The code for the generation of plots are demonstrated in this section as well.

```{r}
#Algorithm to collect and store gated data of the CO2 Variation Experiments
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/Temp & CO2 baseline")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/Temp & CO2 baseline/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 1%")
wd <- getwd()
```

```
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 1%/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 2%")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 2%/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 3%")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 3%/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
```

```
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 4%")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 4%/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 6%")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 6%/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 7%")
wd <- getwd()
```

```
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 7%/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 8%")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 8%/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 9%")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 9%/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
```

```
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 10%")
wd <- getwd()
x_WGA <- c("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 10%/Compensation Controls - WGA")
flow_gating_list <- flow_gating(wd, x_WGA)
#flow_gating_list <- flow_gating(wd)
gs1 <- flow_gating_list[[1]]
gs2 <- flow_gating_list[[2]]
gs3 <- flow_gating_list[[3]]
save_gs(gs1, path = file.path(wd, "gs1"))
save_gs(gs2, path = file.path(wd, "gs2"))
save_gs(gs3, path = file.path(wd, "gs3"))

```

Algorithm to retrieve and to statistically treat the data of CO2 Variation Experiments

```{r }
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/Temp & CO2 baseline")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
CO2_baseline <- table_summary(gs1, gs2, gs3, c("e"))
CO2_baseline_descriptive <- table_descriptive(gs1, gs2, gs3, c("e"))
CO2_baseline_density <- lectin_density_stats(CO2_baseline)
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 1%")
```

```
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
CO2_1 <- table_summary(gs1, gs2, gs3, c("a"))
CO2_1_descriptive <- table_descriptive(gs1, gs2, gs3, c("a"))
base_1 <- F_T_Power_test(CO2_baseline, CO2_1, c("a"))
CO2_1_density <- lectin_density_stats(CO2_1)
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 2%")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
CO2_2 <- table_summary(gs1, gs2, gs3, c("b"))
CO2_2_descriptive <- table_descriptive(gs1, gs2, gs3, c("b"))
base_2 <- F_T_Power_test(CO2_baseline, CO2_2, c("b"))
CO2_2_density <- lectin_density_stats(CO2_2)
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 3%")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
CO2_3 <- table_summary(gs1, gs2, gs3, c("c"))
CO2_3_descriptive <- table_descriptive(gs1, gs2, gs3, c("c"))
base_3 <- F_T_Power_test(CO2_baseline, CO2_3, c("c"))
CO2_3_density <- lectin_density_stats(CO2_3)
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 4%")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
```

```
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
CO2_4 <- table_summary(gs1, gs2, gs3, c("d"))
CO2_4_descriptive <- table_descriptive(gs1, gs2, gs3, c("d"))
base_4 <- F_T_Power_test(CO2_baseline, CO2_4, c("d"))
CO2_4_density <- lectin_density_stats(CO2_4)

setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 6%")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
CO2_6 <- table_summary(gs1, gs2, gs3, c("f"))
CO2_6_descriptive <- table_descriptive(gs1, gs2, gs3, c("f"))
base_6 <- F_T_Power_test(CO2_baseline, CO2_6, c("f"))
CO2_6_density <- lectin_density_stats(CO2_6)
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 7%")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
gs3 <- load_gs(file.path(wd, "gs3"))
CO2_7 <- table_summary(gs1, gs2, gs3, c("g"))
CO2_7_descriptive <- table_descriptive(gs1, gs2, gs3, c("g"))
base_7 <- F_T_Power_test(CO2_baseline, CO2_7, c("g"))
CO2_7_density <- lectin_density_stats(CO2_7)
setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 8%")
wd <- getwd()
gs1 <- load_gs(file.path(wd, "gs1"))
gs2 <- load_gs(file.path(wd, "gs2"))
```

```
gs3 <- load_gs(file.path(wd, "gs3"))

CO2_8 <- table_summary(gs1, gs2, gs3, c("h"))

CO2_8_descriptive <- table_descriptive(gs1, gs2, gs3, c("h"))

base_8 <- F_T_Power_test(CO2_baseline, CO2_8, c("h"))

CO2_8_density <- lectin_density_stats(CO2_8)

setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 9%")

wd <- getwd()

gs1 <- load_gs(file.path(wd, "gs1"))

gs2 <- load_gs(file.path(wd, "gs2"))

gs3 <- load_gs(file.path(wd, "gs3"))

CO2_9 <- table_summary(gs1, gs2, gs3, c("i"))

CO2_9_descriptive <- table_descriptive(gs1, gs2, gs3, c("i"))

base_9 <- F_T_Power_test(CO2_baseline, CO2_9, c("i"))

CO2_9_density <- lectin_density_stats(CO2_9)

setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II/CO2 10%")

wd <- getwd()

gs1 <- load_gs(file.path(wd, "gs1"))

gs2 <- load_gs(file.path(wd, "gs2"))

gs3 <- load_gs(file.path(wd, "gs3"))

CO2_10 <- table_summary(gs1, gs2, gs3, c("j"))

CO2_10_descriptive <- table_descriptive(gs1, gs2, gs3, c("j"))

base_10 <- F_T_Power_test(CO2_baseline, CO2_10, c("j"))

CO2_10_density <- lectin_density_stats(CO2_10)

CO2_global_descriptive_df <- rbind(CO2_baseline_descriptive, CO2_1_descriptive,
CO2_2_descriptive, CO2_3_descriptive, CO2_4_descriptive, CO2_6_descriptive,
CO2_7_descriptive, CO2_8_descriptive, CO2_9_descriptive, CO2_10_descriptive)

rm(CO2_baseline_descriptive, CO2_1_descriptive, CO2_2_descriptive, CO2_3_descriptive,
CO2_4_descriptive, CO2_6_descriptive, CO2_7_descriptive, CO2_8_descriptive,
CO2_9_descriptive, CO2_10_descriptive)

CO2_global_lectinvariation_df <- rbind(CO2_baseline, CO2_1, CO2_2, CO2_3, CO2_4,
CO2_6, CO2_7, CO2_8, CO2_9, CO2_10)
```

```
rm(CO2_baseline, CO2_1, CO2_2, CO2_3, CO2_4, CO2_6, CO2_7, CO2_8, CO2_9, CO2_10)

CO2_global_F_T_df <- rbind(base_1, base_2, base_3, base_4, base_6, base_7, base_8,
base_9, base_10)

rm(base_1, base_2, base_3, base_4, base_6, base_7, base_8, base_9, base_10)

CO2_global_density_df <- rbind(CO2_baseline_density, CO2_1_density, CO2_2_density,
CO2_3_density, CO2_4_density, CO2_6_density, CO2_7_density, CO2_8_density,
CO2_9_density, CO2_10_density)

rm(CO2_baseline_density, CO2_1_density, CO2_2_density, CO2_3_density, CO2_4_density,
CO2_6_density, CO2_7_density, CO2_8_density, CO2_9_density, CO2_10_density)

ratio_matrix_a <- as.data.frame(matrix(c(rep("area", times =
nrow(CO2_global_ratios_a_df))), nrow = nrow(CO2_global_ratios_a_df), ncol = 1),
stringsAsFactors = FALSE)

ratio_matrix_h <- as.data.frame(matrix(c(rep("height", times =
nrow(CO2_global_ratios_h_df))), nrow = nrow(CO2_global_ratios_h_df), ncol = 1),
stringsAsFactors = FALSE)

ratio_matrix_w <- as.data.frame(matrix(c(rep("width", times =
nrow(CO2_global_ratios_w_df))), nrow = nrow(CO2_global_ratios_w_df), ncol = 1),
stringsAsFactors = FALSE)

CO2_global_ratios_df <- rbind(ratio_matrix_a, ratio_matrix_h, ratio_matrix_w)

CO2_global_ratios_df <- cbind(rbind(CO2_global_ratios_a_df, CO2_global_ratios_h_df,
CO2_global_ratios_w_df), CO2_global_ratios_df)

rm(ratio_matrix_a, ratio_matrix_h, ratio_matrix_w, CO2_global_ratios_a_df,
CO2_global_ratios_h_df, CO2_global_ratios_w_df)

colnames(CO2_global_ratios_df) <- c("Sample_Type", "Comp_type", "Ratio1_Mean",
"Ratio1_SD", "Sample_size1", "Ratio2_Mean", "Ratio2_SD","Sample_size2", "Fp_value",
"Level_of_sig_F", "F_test_conclusion", "Tp_value", "Level_of_sig_T", "Power",
"Ascending_order", "Lectin", "Dimension")

colnames(CO2_global_F_T_df) <- c("Sample_Type", "Channels", "Subpopulation", "Lectin",
"Mean", "SD", "Sample_Size","Fp_value", "F_significance", "F_test_conclusion", "Tp_value",
"T_test_significance", "Power")

colnames(CO2_global_lectinvariation_df) <- c("Sample_Type", "Channels", "Mean",
"Mean_SD", "CV_perc", "Subpopulation", "Sample_Size","Viability_perc",
"Viability_SD_perc", "Lectin")
```

```{r}
```

```{r}
library(readxl)

library(gridExtra)
```

#Read in the excel spreadsheet into R

setwd("~/Dropbox/PhD Project/PhD Project/CO2 Variation II")

pH_CO2 <- read_excel("pH.xlsx")

pH_CO2_df <- as.data.frame(pH_CO2, stringsAsFactors = FALSE)

```

**Viability and pH Plots**

```{r}

**#Viability across nutrient variation (line plot of individual lectin curves)**

viability_plot <- mutate(CO2_global_lectinvariation_df, Sample_Type = factor(Sample_Type, levels = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"))) %>%

  ggplot(aes(Sample_Type, Viability_perc)) +

  geom_smooth(aes(group = Lectin, color = Lectin), size = 1.5, se = FALSE) +

  scale_colour_manual(name = "Lectin",  values = c("#980043","#7a0177", "#08519c", "#006d2c", "#7fcdbb",  "#ff7f00", "#993404")) +

   labs(x = "Level of carbon dioxide (%)" , y = "Viability (%)", title = NULL ) +

   scale_x_discrete(expand = c(0,0), breaks = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"), labels = c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +

  theme_classic() +

  theme(

    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),

    legend.text = element_text(size = 15),

    legend.title = element_text(size = 15, face = "bold"),

    legend.box.background = element_blank(),

    legend.justification = "center",

    legend.position = "right",

    axis.title = element_text(size = 15),

    strip.text = element_text(size = 15),

    strip.background = element_rect(fill = "grey90")

   )

**#pH line plot across nutrient variation**

```
pH_plot <- ggplot(pH_CO2_df, aes(Sample_Type, pH)) +
  geom_point(aes(colour = Replicate), size = 3, alpha = 0.60) +
  geom_smooth(size = 1.5) +
  labs(x = "Level of carbon dioxide (%)" , y = "pH", title = NULL) +
  scale_colour_brewer(palette = "Set1", name = "Replicate") +
  scale_x_continuous(expand = c(0,0), breaks = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10), labels = c("-4", "-
3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
  scale_y_continuous(expand = c(0,0)) +
  theme_classic() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "right",
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90")
)
ggplot(media_global_lectinvariation_df, aes(Sample_Type, Viability_perc)) +
  geom_point(alpha = 0.0) +
 geom_smooth(data = viabilityPNA_df, aes(colour = "A"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 8), se = FALSE) +
  geom_smooth(data = viabilityAAL_df, aes(colour = "B"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 7), se = FALSE) +
  geom_smooth(data = viabilityMALII_df, aes(colour = "C"), method = "lm", size = 1.5,
formula = y ~ splines::bs(x, 8), se = FALSE) +
  geom_smooth(data = viabilityLECB_df, aes(colour = "D"), method = "lm", size = 1.5, formula
= y ~ splines::bs(x, 8), se = FALSE) +
```

```
  geom_smooth(data = viabilityLECA_df, aes(colour = "E"), method = "lm", size = 1.5, formula = y ~ splines::bs(x, 8), se = FALSE) +

  geom_smooth(data = viabilityAAL2_df, aes(colour = "F"), method = "lm", size = 1.5, formula = y ~ splines::bs(x, 8), se = FALSE) +

  geom_smooth(data = viabilityWGA_df, aes(colour = "G"), method = "lm", size = 1.5, formula = y ~ splines::bs(x, 6), se = FALSE) +

geom_vline(aes(xintercept = c(6.95)), color = "red", linetype = "dashed", size = 1) +

   labs(x = expression(paste("Lectin concentration (", mu, "g/mL)")) , y = "Viability (%)", title = NULL ) +

  scale_colour_manual(name = "Lectin",  values = c("#ff7f00", "#980043", "#7fcdbb", "#006d2c", "#08519c", "#7a0177", "#993404"), breaks = c("A", "B", "C", "D", "E", "F", "G"), labels = c("PNA", "AAL", "MALII", "LECB", "LECA", "AAL-2", "WGA")) +

scale_x_discrete(expand = c(0,0)) +

  theme_classic() +

  theme(

    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),

    legend.text = element_text(size = 15),

    legend.title = element_text(size = 15, face = "bold"),

    legend.box.background = element_blank(),

    legend.justification = "center",

    legend.position = "right",

    axis.title = element_text(size = 15),

    strip.text = element_text(size = 15),

    strip.background = element_rect(fill = "grey90"),

   )
```

Facetted plots with all lectins - Descriptive Analysis

```{r}
```

#FSC-A

#AAL, LECB, PNA, LECA, AAL-2, WGA, MAL II

```
viability_plot <- mutate(CO2_global_lectinvariation_df, Sample_Type = factor(Sample_Type, levels = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"))) %>%
```

```
  ggplot(aes(Sample_Type, Viability_perc)) +
  geom_smooth(aes(group = Lectin, color = Lectin), size = 1.5, se = FALSE) +
  scale_colour_manual(name = "Lectin",  values = c("#980043","#7a0177", "#08519c",
"#006d2c", "#7fcdbb",  "#ff7f00", "#993404")) +
   labs(x = "Level of carbon dioxide (%)" , y = "Viability (%)", title = NULL ) +
   scale_x_discrete(expand = c(0,0), breaks = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),
labels = c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
filter(CO2_global_lectinvariation_df, Channels == "FSC-A", Subpopulation != "Dead +
Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S",
"Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
"MAL II"))) %>%
  mutate(Sample_Type = factor(Sample_Type, levels = c("1", "2", "3", "4", "5", "6", "7", "8",
"9", "10"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
  geom_point() +
  geom_line(size = 1) +
  theme_classic() +
  labs(x = "Level of carbon dioxide (%)" , y = "FSC-A (linear scale)", title = NULL ) +
  scale_colour_brewer(palette = "Dark2", name = "Subpopulation") +
 scale_x_discrete(expand = c(0,0), breaks = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),
labels = c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
  facet_grid(.~ Lectin) +
  theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
```

```
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
   )
#SSC-A
#All populations
p1 <- filter(CO2_global_lectinvariation_df, Channels == "SSC-A", Subpopulation != "Dead +
Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S",
"Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
"MAL II"))) %>%
  mutate(Sample_Type = factor(Sample_Type, levels = c("1", "2", "3", "4", "5", "6", "7", "8",
"9", "10"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
  geom_point() +
  geom_line(size = 1) +
  theme_classic() +
  labs(x = NULL , y = "SSC-A (linear scale)", title = NULL ) +
  scale_colour_brewer(palette = "Dark2", name = "Subpopulation", guide = FALSE) +
  scale_x_discrete(expand = c(0,0), breaks = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),
labels = c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
  facet_grid(.~ Lectin) +
  theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
```

```
    legend.justification = "center",
    legend.position = "bottom",
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
   )
#SSC-A
#DNA cycle populations
p2 <- filter(CO2_global_lectinvariation_df, Channels == "SSC-A", Subpopulation != "Dead + Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S", "Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))) %>%
  mutate(Sample_Type = factor(Sample_Type, levels = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
  geom_point() +
  geom_line(size = 1) +
  theme_classic() +
  labs(x = "Level of carbon dioxide (%)"  , y = "SSC-A (linear scale)", title = NULL ) +
  scale_colour_brewer(palette = "Dark2", name = "Subpopulation") +
   scale_x_discrete(expand = c(0,0), breaks = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"), labels = c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
  scale_y_continuous(limits = c(15000, 50000)) +
  facet_grid(.~ Lectin) +
  theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
```

```
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
   )
grid.arrange(p1, p2, nrow = 2)
#LECTIN-A
#All populations
p1 <- filter(CO2_global_lectinvariation_df, Channels == "LECTIN-A", Subpopulation != "Dead
+ Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead", "Apoptotic", "G2/M", "S",
"Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
"MAL II"))) %>%
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
  geom_point() +
  geom_line(size = 1) +
  theme_classic() +
  labs(x = NULL , y = "LECTIN-A (linear scale)", title = NULL ) +
  scale_colour_brewer(palette = "Dark2", name = "Subpopulation", guide = FALSE) +
scale_x_discrete(expand = c(0,0), breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels
= c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
  facet_grid(.~ Lectin) +
  theme_bw() +
  theme(
```

```
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
  )
#LECTIN-A
#1b9e77 Dead
#d95f02 Apoptotic
#7570b3 G2/M
#e7298a S
#66a61e Go/G1
#DNA cycle populations
df1 <- filter(CO2_global_lectinvariation_df, Channels == "LECTIN-A", Subpopulation ==
"Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead","Apoptotic", "G2/M", "S",
"Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
"MAL II")))
p2 <- filter(CO2_global_lectinvariation_df, Channels == "LECTIN-A", Subpopulation != "Dead
+ Apoptotic") %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("Dead","Apoptotic", "G2/M", "S",
"Go/G1"))) %>%
  mutate(Lectin = factor(Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA",
"MAL II"))) %>%
```

```
ggplot(aes(Sample_Type, Mean, group = Subpopulation, colour = Subpopulation)) +
  geom_point() +
  geom_line(size = 1) +
  theme_classic() +
  labs(x = "Level of carbon dioxide (%)", y = "LECTIN-A (linear scale)", title = NULL ) +
  scale_colour_brewer(palette = "Dark2", name = "Subpopulation") +
  geom_line(data = df1, colour = "white", size = 2) +
   geom_point(data = df1, colour = "white", size = 2) +
scale_x_discrete(expand = c(0,0), breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels
= c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
  scale_y_continuous(limits = c(30, 350)) +
  facet_grid(.~ Lectin) +
  theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
   )
grid.arrange(p1, p2, nrow = 2)
```

Lectin Inferential Analysis

```{r}

Lectin_A_Subp_G2M_df <- table_manipulation(CO2_global_descriptive_df,
CO2_global_F_T_df, c("LECTIN_A"), c("LECTIN-A"), c("G2/M"), c("CO2"))

Lectin_A_Subp_S_df <- table_manipulation(CO2_global_descriptive_df, CO2_global_F_T_df,
c("LECTIN_A"), c("LECTIN-A"), c("S"), c("CO2"))

Lectin_A_Subp_GoG1_df <- table_manipulation(CO2_global_descriptive_df,
CO2_global_F_T_df, c("LECTIN_A"), c("LECTIN-A"), c("Go/G1"), c("CO2"))

Lectin_A_df <- rbind(Lectin_A_Subp_G2M_df, Lectin_A_Subp_S_df,
Lectin_A_Subp_GoG1_df)

Lectin_A_df$Lectin_face <- factor(Lectin_A_df$Lectin, levels = c("AAL", "LEC B", "PNA", "LEC
A", "AAL-2", "WGA", "MAL II"))

Lectin_A_df$Subpopulation_face <- factor(Lectin_A_df$Subpopulation, levels = c("G2/M",
"S", "Go/G1", "Apoptotic","Dead"))

CO2_global_descriptive_df$Lectin_face <- factor(CO2_global_descriptive_df$Lectin, levels =
c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))

CO2_global_descriptive_df$Subpopulation_face <-
factor(CO2_global_descriptive_df$Subpopulation, levels = c("G2/M", "S", "Go/G1",
"Apoptotic","Dead"))

CO2_global_lectinvariation_df$Lectin_face <- factor(CO2_global_lectinvariation_df$Lectin,
levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))

**#Lectin_A_df$Sample_Type_face <- factor(Lectin_A_df$Sample_Type, levels = c("1", "2",
"3", "4", "5", "6", "7", "8", "9", "10"))**

**#CO2_global_descriptive_df$Sample_Type_face <-
factor(CO2_global_descriptive_df$Sample_Type, levels = c("1", "2", "3", "4", "5", "6", "7",
"8", "9", "10"))**

**#set fill and colour manual**

**#d95f02 highly significant**

**#1b9e77 not significant**

**#7570b3 trend towards significance**

**#e7298a very highly significant**

**#66a61e significant**

**# plot facetted by lectin and subpopulation**

df1 <- filter(CO2_global_descriptive_df, Sample_Type == 'e', Subpopulation %in% c("Go/G1",
"S", "G2/M"))

filter(Lectin_A_df, Subpopulation %in% c("Go/G1", "S", "G2/M")) %>%
```

```
ggplot(aes(Sample_Type, LECTIN_A)) +
  geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
  geom_boxplot(data = filter(CO2_global_descriptive_df, Sample_Type == 'e', Subpopulation %in% c("Go/G1", "S", "G2/M")), aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20, outlier.shape = NA) +
  facet_grid(Subpopulation_face~ Lectin_face) +
  labs(x = "Level of carbon dioxide (%)", y = "LECTIN-A (linear scale)", title = NULL) +
  scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels = c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
   scale_fill_manual(name = "Level of Statistical Significance", values = c("#d95f02", "#1b9e77", "#e7298a")) +
   scale_colour_manual(values = c("#d95f02", "#1b9e77", "#e7298a"), guide = FALSE) +
   scale_y_continuous(expand = c(0,0)) +
 coord_cartesian(ylim = c(-80, 700)) +
 theme_bw() +
  theme(
   plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
   legend.text = element_text(size = 15),
   legend.title = element_text(size = 15, face = "bold"),
   legend.box.background = element_blank(),
   legend.justification = "center",
   legend.position = "bottom",
   axis.title = element_text(size = 15),
   strip.text = element_text(size = 15),
   strip.background = element_rect(fill = "grey90"),
   panel.grid = element_blank(),
   panel.spacing = unit(0.75, "lines")
  )
df1 <-  filter(CO2_global_lectinvariation_df, Channels == "LECTIN-A", Subpopulation == "Go/G1") %>%
ggplot(aes(Sample_Type, Mean, colour = "#66a61e")) +
```

```
  geom_point() +
  geom_line(size = 1) +
  labs(x = NULL , y = "LECTIN-A (log scale)", title = NULL ) +
  scale_colour_brewer(palette = "Dark2", name = "Subpopulation", guide = FALSE) +
  scale_x_discrete(expand = c(0,0), breaks = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),
labels = c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
  facet_grid(.~ Lectin) +
  theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
   )
plot_line_box <-  filter(Lectin_A_df, Subpopulation == "Go/G1")  %>%
ggplot(aes(Sample_Type, LECTIN_A)) +
  geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
  geom_boxplot(data = filter(CO2_global_descriptive_df, Sample_Type == 'e', Subpopulation
== "Go/G1"), aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20, outlier.shape = NA) +
  #geom_point(data = filter(CO2_global_lectinvariation_df, Channels == "LECTIN-A",
Subpopulation == "Go/G1"), aes(Sample_Type, Mean, group = Subpopulation), colour =
"black", size = 0.6) +
```

```
  geom_line(data = filter(CO2_global_lectinvariation_df, Channels == "LECTIN-A",
Subpopulation == "Go/G1"), aes(Sample_Type, Mean, group = Subpopulation), colour =
"black", size = 0.6) +
   facet_grid(~ Lectin_face) +
  labs(x = "Level of carbon dioxide (%)", y = "LECTIN-A (linear scale)", title = NULL) +
   scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels = c("-4", "-3",
"-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
    scale_fill_manual(name = "Level of Statistical Significance", values = c("#d95f02",
"#1b9e77", "#e7298a")) +
     scale_colour_manual(values = c("#d95f02", "#1b9e77", "#e7298a"), guide = FALSE) +
     scale_y_continuous(expand = c(0,0)) +
coord_cartesian(ylim = c(-50, 460)) +
 theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
   )
df1 <- filter(CO2_global_descriptive_df, Sample_Type == 'e', Subpopulation == "G2/M")
p_G2M  <- filter(Lectin_A_df, Subpopulation == "G2/M") %>%
  ggplot(aes(Sample_Type, LECTIN_A)) +
  geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
```

```
  geom_boxplot(data = df1, aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20,
outlier.shape = NA) +
  facet_grid(Subpopulation ~ Lectin_face) +
  labs(x = "Level of carbon dioxide (%)", y = "LECTIN-A (log scale)", title = NULL) +
  scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels = c("-4", "-3",
"-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
    scale_fill_manual(name = "Level of Statistical Significance", values = c("#1b9e77",
"#e7298a")) +
    scale_colour_manual(values = c("#1b9e77", "#e7298a"), guide = FALSE) +
    scale_y_continuous(expand = c(0,0)) +
coord_cartesian(ylim = c(-50, 460)) +
 theme_bw() +
  theme(
   plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
   legend.text = element_text(size = 15),
   legend.title = element_text(size = 15, face = "bold"),
   legend.box.background = element_blank(),
   legend.justification = "center",
   legend.position = "bottom",
   axis.title = element_text(size = 15),
   strip.text = element_text(size = 15),
   strip.background = element_rect(fill = "grey90"),
   panel.grid = element_blank(),
   panel.spacing = unit(0.75, "lines")
  )
#d95f02 highly significant
#1b9e77 not significant
#7570b3 trend towards significance
#e7298a very highly significant
#66a61e significant
```

```
df1 <- filter(CO2_global_descriptive_df, Sample_Type == 'e', Subpopulation == "S")
p_S <- filter(Lectin_A_df, Subpopulation == "S") %>%
  ggplot(aes(Sample_Type, LECTIN_A)) +
  geom_boxplot(aes(fill = T_test_significance), size = 0.20, outlier.shape = NA) +
  geom_boxplot(data = df1, aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.20,
outlier.shape = NA) +
  facet_grid(Subpopulation ~ Lectin_face) +
  labs(x = "Level of carbon dioxide (%)", y = "LECTIN-A (log scale)", title = NULL) +
  scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels = c("-4", "-3",
"-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
    scale_fill_manual(name = "Level of Statistical Significance", values = c("#1b9e77",
"#e7298a")) +
    scale_colour_manual(values = c("#1b9e77", "#e7298a"), guide = FALSE) +
    scale_y_continuous(expand = c(0,0)) +
 coord_cartesian(ylim = c(-65, 600)) +
 theme_bw() +
  theme(
   plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
   legend.text = element_text(size = 15),
   legend.title = element_text(size = 15, face = "bold"),
   legend.box.background = element_blank(),
   legend.justification = "center",
   legend.position = "bottom",
   axis.title = element_text(size = 15),
   strip.text = element_text(size = 15),
   strip.background = element_rect(fill = "grey90"),
   panel.grid = element_blank(),
   panel.spacing = unit(0.75, "lines")
  )
```

#d95f02 highly significant

#1b9e77 not significant

#7570b3 trend towards significance

#e7298a very highly significant

#66a61e significant

```
df1 <- filter(CO2_global_descriptive_df, Sample_Type == 'e', Subpopulation == "Go/G1")
p_Go <- filter(Lectin_A_df, Subpopulation == "Go/G1") %>%
  ggplot(aes(Sample_Type, LECTIN_A)) +
  geom_boxplot(aes(fill = T_test_significance), size = 0.2, outlier.shape = NA) +
  geom_boxplot(data = df1, aes(Sample_Type, LECTIN_A), fill = "grey", size = 0.2,
outlier.shape = NA) +
  facet_grid(Subpopulation ~ Lectin_face) +
  labs(x = "Level of carbon dioxide (%)", y = "LECTIN-A (log scale)", title = NULL) +
  scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels = c("-4", "-3",
"-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
  scale_fill_manual(name = "Level of Statistical Significance", values = c("#1b9e77",
"#66a61e", "#e7298a")) +
  scale_colour_manual(values = c("#1b9e77", "#66a61e", "#e7298a"), guide = FALSE) +
  scale_y_continuous(expand = c(0,0)) +
  coord_cartesian(ylim = c(-50, 460)) +
  theme_bw() +
  theme(
   plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
   legend.text = element_text(size = 15),
   legend.title = element_text(size = 15, face = "bold"),
   legend.box.background = element_blank(),
   legend.justification = "center",
   legend.position = "bottom",
   axis.title = element_text(size = 15),
   strip.text = element_text(size = 15),
```

```
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
   )
```

Lectin Power Analysis

```{r}
**#fd8d3c G2/M**
**#f03b20 S**
**#bd0026 Go/G1**
CO2_global_F_T_df$Lectin_face <- factor(CO2_global_F_T_df$Lectin, levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))
filter(CO2_global_F_T_df, Channels == "LECTIN-A", Subpopulation %in% c("G2/M", "S", "Go/G1")) %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("G2/M", "S", "Go/G1"))) %>%
  mutate(Power = Power * 100) %>%
   ggplot(aes(Sample_Type, Power, fill = Subpopulation)) +
   geom_bar(stat = "identity", colour = NA) +
   facet_grid(Subpopulation ~ Lectin_face) +
   labs(x = "Level of carbon dioxide (%)", y = "Power (%)", title = NULL) +
  #scale_fill_brewer(palette = "RdBu", guide = FALSE) +
   scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels = c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
   scale_fill_manual(values = c("#fd8d3c", "#f03b20", "#bd0026"), guide = FALSE) +
   scale_colour_manual(values = c("#fd8d3c", "#f03b20", "#bd0026"), guide = FALSE) +
   scale_y_continuous(expand = c(0,0), limits = c(0,100)) +
   theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
```

```
3780      legend.title = element_text(size = 15, face = "bold"),
3781      legend.box.background = element_blank(),
3782      legend.justification = "center",
3783      legend.position = "bottom",
3784      axis.title = element_text(size = 15),
3785      strip.text = element_text(size = 15),
3786      strip.background = element_rect(fill = "grey90"),
3787      panel.grid = element_blank(),
3788      panel.spacing = unit(0.75, "lines")
3789     )
3790  bar_plot <- filter(CO2_global_F_T_df, Subpopulation =='Go/G1', Channels == "LECTIN-A")
3791  %>%
3792    mutate(Power = Power * 100) %>%
3793    ggplot(aes(Sample_Type, Power)) +
3794    geom_rect(aes(xmin = 0.4, xmax = 4.5, ymin = 0, ymax = Inf), fill = "#bd0026", alpha =
3795  0.025) +
3796    geom_rect(aes(xmin = 4.5, xmax = Inf, ymin = 0, ymax = Inf), fill = "#bd0026", alpha = 0.07)
3797  +
3798    geom_bar(stat = "identity" , colour = NA, fill = "#bd0026") +
3799    geom_text(data = filter(CO2_global_F_T_df, Subpopulation =='Go/G1', Channels ==
3800  "LECTIN-A") %>%
3801    mutate(Power = Power * 100) %>% mutate_if(is.numeric, round, 0), aes(Sample_Type,
3802  Power, label = Power), position = position_dodge(width = 0.8), size = 4, vjust = -0.5) +
3803    facet_grid(.~ Lectin_face) +
3804    labs(x = NULL, y = "Power (%)", title = NULL) +
3805    scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels = c("-4", "-3",
3806  "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
3807   scale_y_continuous(expand = c(0,0), limits = c(0,110), breaks = c(0, 25, 50, 75, 100)) +
3808    theme_bw() +
3809   theme(
3810    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
```

```
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank(),
    panel.spacing = unit(0.75, "lines")
   )
grid.arrange(bar_plot, plot_line_box, nrow = 2)
```

Sample Size

```{r}
**#41b6c4**

**#2c7fb8**

**#253494**

filter(CO2_global_F_T_df, Channels == "LECTIN-A", Subpopulation %in% c("G2/M", "S", "Go/G1")) %>%
  mutate(Subpopulation = factor(Subpopulation, levels = c("G2/M", "S", "Go/G1"))) %>%
   ggplot(aes(Sample_Type, Sample_Size, fill = Subpopulation)) +
   geom_bar(stat = "identity", colour = NA) +
   facet_grid(Subpopulation ~ Lectin_face) +
   labs(x = "Level of carbon dioxide (%)", y = "Sample size (number of cells)", title = NULL) +
   scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels = c("-4", "-3", "-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
   scale_fill_manual(values = c("#41b6c4", "#2c7fb8", "#253494"), guide = FALSE) +
   scale_colour_manual(values = c("#41b6c4", "#2c7fb8", "#253494"), guide = FALSE) +
   scale_y_continuous(expand = c(0,0)) +
```

```
 theme_bw() +
 theme(
  plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
  legend.text = element_text(size = 15),
  legend.title = element_text(size = 15, face = "bold"),
  legend.box.background = element_blank(),
  legend.justification = "center",
  legend.position = "bottom",
  axis.title = element_text(size = 15),
  strip.text = element_text(size = 15),
  strip.background = element_rect(fill = "grey90"),
  panel.grid = element_blank(),
  panel.spacing = unit(0.75, "lines")
 )
```

Analysis of Relative Lectin signal density

```{r}
library(gridExtra)

CO2_global_lectinvariation_df$Subpopulation_face <-
factor(CO2_global_lectinvariation_df$Subpopulation, levels = c("Go/G1", "S", "G2/M",
"Apoptotic","Dead"))

CO2_global_lectinvariation_df$Lectin_face <- factor(CO2_global_lectinvariation_df$Lectin,
levels = c("AAL", "LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))

#b2e2e2 Go/G1

#66c2a4 S

#238b45 G2/M

p1 <- filter(CO2_global_lectinvariation_df, Channels %in% c("Area_ratio"),
Subpopulation_face %in% c("G2/M", "S", "Go/G1"), Lectin_face %in% c("AAL", "LEC B",
"PNA", "LEC A")) %>%

  ggplot(aes(Sample_Type, Mean, fill = Subpopulation_face, ymin = Mean - Mean_SD, ymax =
Mean + Mean_SD, group = Subpopulation_face)) +
```

```
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(size = 0.15, position = "dodge") +
   facet_grid(.~ Lectin_face) +
   labs(x = "Level of carbon dioxide (%)", y = "LECTIN-A/FSC-A (linear scale)", title = NULL) +
   scale_fill_manual(values = c("#b2e2e2", "#66c2a4", "#238b45"), guide = FALSE) +
   scale_y_continuous(expand = c(0,0), limits = c(0,0.0025)) +
 scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels = c("-4", "-3",
"-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
   theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.text.x = element_text(size = 10, colour = "black"),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank()
  )
p2 <- filter(CO2_global_lectinvariation_df, Channels %in% c("Area_ratio"),
Subpopulation_face %in% c("G2/M", "S", "Go/G1"), Lectin_face %in% c("AAL-2", "WGA",
"MAL II")) %>%
  ggplot(aes(Sample_Type, Mean, fill = Subpopulation_face, ymin = Mean - Mean_SD, ymax =
Mean + Mean_SD, group = Subpopulation_face)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(size = 0.15, position = "dodge") +
   facet_grid(.~ Lectin_face) +
   labs(x = "Level of carbon dioxide (%)", y = "LECTIN-A/FSC-A (linear scale)", title = NULL) +
```

```
  scale_fill_manual(values = c("#b2e2e2", "#66c2a4", "#238b45"), name = "Subpopulation") +
   scale_y_continuous(expand = c(0,0), limits = c(0,0.0035)) +
 scale_x_discrete(breaks = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"), labels = c("-4", "-3",
"-2", "-1", "0", "+1", "+2", "+3", "+4", "+5")) +
   theme_bw() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "right",
    axis.text.x = element_text(size = 10, colour = "black"),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank()
  )
grid.arrange(p1, p2, nrow = 2)
```
```
```
Averaging power
```{r}
**#By Lectin**
**#LECA – blue or #08519c, #3182bd, #6baed6, #9ecae1, #c6dbef, #eff3ff (single hue)**
**#LECB – green or #006d2c, #2ca25f, #66c2a4, #99d8c9, #ccece6, #edf8fb (multi-hue)**
**#AAL2 – violet or #7a0177, #c51b8a, #f768a1, #fa9fb5, #fcc5c0, #feebe2 (multi-hue)**
**#AAL – red or #980043, #dd1c77, #df65b0, #c994c7, #d4b9da, #f1eef6 (multi-hue)**
**#PNA – orange #ff7f00 get shades from alpha levels**
```

```
#WGA – brown or #993404,  #d95f0e, #fe9929, #fed98e, #ffffd4 (multi-hue)
#MALII – #7fcdbb get shades from alpha levels
#Spent medium
media_global_F_T_df$Lectin_face <- factor(media_global_F_T_df$Lectin, levels = c("AAL",
"LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))
df1 <- filter(media_global_F_T_df, Subpopulation == "Go/G1", Channels == "LECTIN-A") %>%
group_by(Lectin_face) %>%
  mutate(Power = Power *100) %>%
  summarise(P_mean = mean(Power))
filter(media_global_F_T_df, Subpopulation == "Go/G1", Channels == "LECTIN-A") %>%
group_by(Lectin_face) %>%
  mutate(Power = Power *100) %>%
  summarise(P_mean = mean(Power)) %>%
ggplot(aes(Lectin_face, P_mean)) +
 geom_bar(aes(colour = NULL, fill = Lectin_face), stat = "identity", alpha = 0.70) +
  geom_text(data = mutate_if(df1, is.numeric, round, 2), aes(Lectin_face, P_mean, label =
P_mean), position = position_dodge(width = 0.8), size = 5, vjust = -0.5) +
   labs(x =  NULL, y = "Power (%)", title = NULL) +
  scale_fill_manual(values = c("#980043", "#006d2c", "#ff7f00", "#08519c", "#7a0177",
"#993404","#7fcdbb"), guide = FALSE) +
  scale_y_continuous(expand = c(0,0), limits = c(0, 100)) +
   theme_classic() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.text.x = element_text(size = 12, colour = "black"),
```

```
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank()
  )
#Temperature
temp_global_F_T_df$Lectin_face <- factor(temp_global_F_T_df$Lectin, levels = c("AAL",
"LEC B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))
df2 <- filter(temp_global_F_T_df, Subpopulation == "Go/G1", Channels == "LECTIN-A") %>%
group_by(Lectin_face) %>%
  mutate(Power = Power *100) %>%
  summarise(P_mean = mean(Power))
filter(temp_global_F_T_df, Subpopulation == "Go/G1", Channels == "LECTIN-A") %>%
group_by(Lectin_face) %>%
  mutate(Power = Power *100) %>%
  summarise(P_mean = mean(Power)) %>%
ggplot(aes(Lectin_face, P_mean)) +
 geom_bar(aes(colour = NULL, fill = Lectin_face), stat = "identity", alpha = 0.70) +
  geom_text(data = mutate_if(df2, is.numeric, round, 2), aes(Lectin_face, P_mean, label =
P_mean), position = position_dodge(width = 0.8), size = 5, vjust = -0.5) +
   labs(x =  NULL, y = "Power (%)", title = NULL) +
  scale_fill_manual(values = c("#980043", "#006d2c", "#ff7f00", "#08519c", "#7a0177",
"#993404","#7fcdbb"), guide = FALSE) +
  scale_y_continuous(expand = c(0,0), limits = c(0, 100)) +
   theme_classic() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
```

```
3992        legend.justification = "center",
3993        legend.position = "bottom",
3994        axis.text.x = element_text(size = 12, colour = "black"),
3995        axis.title = element_text(size = 15),
3996        strip.text = element_text(size = 15),
3997        strip.background = element_rect(fill = "grey90"),
3998        panel.grid = element_blank()
3999      )
4000    #CO2
4001    CO2_global_F_T_df$Lectin_face <- factor(CO2_global_F_T_df$Lectin, levels = c("AAL", "LEC
4002    B", "PNA", "LEC A", "AAL-2", "WGA", "MAL II"))
4003    df3 <- filter(CO2_global_F_T_df, Subpopulation == "Go/G1", Channels == "LECTIN-A") %>%
4004    group_by(Lectin_face) %>%
4005      mutate(Power = Power *100) %>%
4006      summarise(P_mean = mean(Power))
4007    filter(CO2_global_F_T_df, Subpopulation == "Go/G1", Channels == "LECTIN-A") %>%
4008    group_by(Lectin_face) %>%
4009      mutate(Power = Power *100) %>%
4010      summarise(P_mean = mean(Power)) %>%
4011    ggplot(aes(Lectin_face, P_mean)) +
4012     geom_bar(aes(colour = NULL, fill = Lectin_face), stat = "identity", alpha = 0.70) +
4013      geom_text(data = mutate_if(df3, is.numeric, round, 2), aes(Lectin_face, P_mean, label =
4014    P_mean), position = position_dodge(width = 0.8), size = 5, vjust = -0.5) +
4015       labs(x =  NULL, y = "Power (%)", title = NULL) +
4016      scale_fill_manual(values = c("#980043", "#006d2c", "#ff7f00", "#08519c", "#7a0177",
4017    "#993404","#7fcdbb"), guide = FALSE) +
4018      scale_y_continuous(expand = c(0,0), limits = c(0, 100)) +
4019       theme_classic() +
4020      theme(
4021        plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
```

```
    legend.text = element_text(size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.box.background = element_blank(),
    legend.justification = "center",
    legend.position = "bottom",
    axis.text.x = element_text(size = 12, colour = "black"),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 15),
    strip.background = element_rect(fill = "grey90"),
    panel.grid = element_blank()
  )
#By cell culture parameter
p_media <- filter(media_global_F_T_df, Subpopulation == "Go/G1", Channels == "LECTIN-A")
%>%
  mutate(Power = Power *100) %>%
  summarise(mean(Power))
p_temp <- filter(temp_global_F_T_df, Subpopulation == "Go/G1", Channels == "LECTIN-A")
%>%
  mutate(Power = Power *100) %>%
  summarise(mean(Power))
p_CO2 <- filter(CO2_global_F_T_df, Subpopulation == "Go/G1", Channels == "LECTIN-A")
%>%
  mutate(Power = Power *100) %>%
  summarise(mean(Power))
power_df <- rbind(p_media, p_temp, p_CO2)
parameter_matrix <- as.data.frame(matrix(c("Spent medium", "Temperature", "Carbon
dioxide")), nrow = 3, ncol = 1, stringsAsFactors = FALSE)
power_df <- cbind(power_df, parameter_matrix)
colnames(power_df) <- c("Power", "Treatment")
ggplot(power_df, aes(Treatment, Power)) +
geom_bar(aes(colour = Treatment, fill = Treatment), stat = "identity") +
```

```
4053  geom_text(data = mutate_if(power_df, is.numeric, round, 2), aes(Treatment, Power, label =
4054  Power), position = position_dodge(width = 0.8), size = 8, vjust = -0.5) +
4055    labs(x =  NULL, y = "Power (%)", title = NULL) +
4056    scale_fill_brewer(palette = "Dark2", guide = FALSE) +
4057    scale_colour_brewer(palette = "Dark2", guide = FALSE) +
4058      #scale_x_discrete(breaks = c("Carbon", "b", "c", "d", "e", "f", "g"), labels = c("-3", "-2", "-
4059  1", "0", "+1", "+2", "+3")) +
4060    scale_y_continuous(expand = c(0,0), limits = c(0, 100)) +
4061     theme_classic() +
4062    theme(
4063      plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
4064      legend.text = element_text(size = 15),
4065      legend.title = element_text(size = 15, face = "bold"),
4066      legend.box.background = element_blank(),
4067      legend.justification = "center",
4068      legend.position = "bottom",
4069      axis.text.x = element_text(size = 12, colour = "black"),
4070      axis.title = element_text(size = 15),
4071      strip.text = element_text(size = 15),
4072      strip.background = element_rect(fill = "grey90"),
4073      panel.grid = element_blank()
4074  )
```