Glyco-profiling of the surface of colorectal carcinoma cell lines using lectin probes

Thesis submitted

for degree of

Doctor of Philosophy

By

Donal Monaghan, B.Sc.

Supervised by

Dr. Brendan O' Connor

and

Dr. Dermot Walls

School of Biotechnology

Dublin City University

Ireland

August 2017

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Degree 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: _____

Student No.: 59432183

Date: _____

Acknowledgments

I would like to take this opportunity to thank a few people.

Thank you to my Mam and Dad for the continued support throughout the last four years. This project would not have been possible without your continued support.

Thank you to my brothers, especially Ruairi as he had been through this process before, for the encouragement given over the course of the project.

Thank you to Brendan and Dermot for the opportunity to pursue a PhD and for all their guidance throughout the project, your continued encouragement and open door policy definitely made the process go as smooth as possible.

Thank you to all my lab mates throughout the four years for making the process a little less painful. A special thank you to Jonathan who had to put up with me annoying him for four years. I would not have gotten through the project otherwise.

And lastly thank you to all the other postgrads in the School of Biotechnology. The events and nights out were a welcome distraction from the PhD at times and again made the whole process over the last four years a lot more enjoyable.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 5

Appendix A

Appendix B

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Abbreviations

A	Alanine
AAL	*Aleuria Aurantia* Lectin
ACS	American Chemical Society
APS	Ammonium persulfate
BCA	Bicinchoninic Acid
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Abumin
Con A	Concanavalin A
CRC	Colorectal Cancer
CV	Column Volume
DBA	*Dolichos Biflorus* Agglutinin
dH_2O	Distilled water
DSL	*Datura Stramonium* Lectin
ECL	*Erythrina Cristagalli* Lectin
ELLA	Enzyme Linked Lectin Assay
ER	Endoplasmic Reticulum
FSC	Forward Scatter
G	Glycine
GalNAc	N-acteylgalactosamine
GFP	Green Fluorescent Protein
GlcNAc	N-acetylglucosamine
GNL	*Galanthus Nivalis* Lectin
GSL I	*Griffonia Simplicifolia* Lectin I
GSL II	*Griffonia Simplicifolia* Lectin II
HPA	*Helix Pomatia* Agglutinin

IMAC	Immobilised Metal Affinity Chromatography
IMS	Industrial Methylated Spirits
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
ISSC	Irish Separation Science Cluster
K	Lysine
LB	Luria Bertani Broth
LCA	*Lens Culinaris* Agglutinin
MAL I	*Maackia Amurensis* Lectin I
MAL II	*Maackia Amurensis* Lectin II
MALDI-TOF	Matrix Assisted Laser Desorption/Ionisation Time-of-Flight
MCS	Multiple Cloning Site
MWCO	Molecular weight cut off
NeuNAc	N-Acetylneuraminic Acid / sialic acid
Ni-NTA	Nickel-nitrolotriacetic Acid
NPL	*Narcissus Pseudonarcissus* Lectin
ORF	Open Reading Frame
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PNA	Peanut Agglutinin
PNGaseF	Peptide:N-glycosidase F
PTM	Post Translation Modifications
PVA	Polyvinyl Alcohol
RCA	*Ricinus Communis* Agglutinin
rpm	Rotations Per Minute
S	Serine
SBA	Soybean Agglutinin

SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SNA	*Sambucus Nigra* Lectin
SOB	Super Optimal Broth
SSC	Side Scatter
ST3 GAL	Sialyltransferase β-Galactoside
T	Threonine
TAE	Tris Acetate EDTA
TB	Terrific Broth
TBS	Tris Buffered Saline
TMB	3,3,5,5-Tetramethylbensidine
Tn	Thomsen-Nouvelle
UEA-I	*Ulex Europaeus* Agglutinin
WGA	Wheat Germ agglutinin
WHO	World Health Organisation

Abstract

The study of cell surface glycosylation has increasingly been seen to be an important area in recent years. A cells ability to carry out its specific function hinges on its ability to correctly glycosylate its surface proteins. The cell membrane has a large number of embedded surface glycoproteins. Correct glycosylation is imperative for correct protein folding to facilitate a number of cellular processes, including; cell-cell signalling and cell adhesion. Alterations to these cell surface glycans can have detrimental effects for the cell and lead to a change in certain cellular functions. It is also recognised that changes in cell surface glycosylation are early indicators of the cell becoming stressed or diseased. To date, antibody probes are primarily utilised to interogate the surface of a cell, usually recognising surface proteins. However, this strategy has not worked for the study of surface glycosylation as antibodies are unable to recognise and distinguish between neutral glycans. Therefore, in order to understand the specific changes to the cell surface glycosylation, new technologies, including new probes, need to be explored and expanded to identify possible changes as accurately as possible.

This work aimed to develop and modify a number of recombinant lectin probes that could be used in unison with commercially available eukaryotic lectin probes to interrogate the cell surface glycosylation of two colorectal carcinoma cell lines. In cancer cells, altered glycosylation is believed to be amongst the earliest changes to occur. These changes could also indicate the specific stages of certain cancers and infer if a cancer is metastasising or not. This work also reports on the use of a panel of lectins, in combination with flow cytometry and fluorescent microscopy, to identify glycosylation changes between an SW480 and SW620 cell line. These paired cell lines are from the same patient with SW480 classed as non-metastatic and SW620 classed as metastatic. This work shows a significant change in the levels of sialylation, both in type and abundance, between the two different cell lines. With a decrease in lectin binding for sialic acid, there was a proportional increase in galactophilic lectin binding which highlights the altered glycosylation found on the cell surface of these cell lines.

Chapter 1

Introduction

1.1 Glycobiology

In recent years the study of glycobiology has become increasingly important. Glycobiology is the study of all carbohydrates, or glycans, attached to proteins and other molecules throughout eukaryotic and prokaryotic life. The process of adding specific glycan chains to proteins, or other biomolecules, is called glycosylation and is deemed as one of the most important post translational modifications (PTM). In recent decades the study of glycosylation has increased exponentially. Since the human genome project was completed, the myth of one gene for one protein product was cast aside. The study of proteomics became more important and all PTMs, including glycosylation, needed to be studied in greater detail. Glycan structures have been shown to be pivotal in nearly all cellular functions; cell signalling (Ohtsubo and Marth 2006), cell-cell adhesion (Gu et al. 2012), cell surface transport proteins as well as internal cytoplasmic signalling in the cell. Glycosylation is described as a non-template driven PTM, which in itself is interesting, as protein synthesis itself is fully template driven (Dwek 1998). Due to the role of glycosylation in a variety of important cellular functions, any alteration in glycan formation or glycan attachment can have detrimental effects on the cell. Most receptors and transfer proteins need to be correctly glycosylated to fulfil their specific function. In figure 1.1 the differences between hexose monosaccharides is the positioning of the H and OH groups on certain carbons in the six carbon structure. Using different hexose sugars, eukaryotes build up these glycans to suit their specific needs.

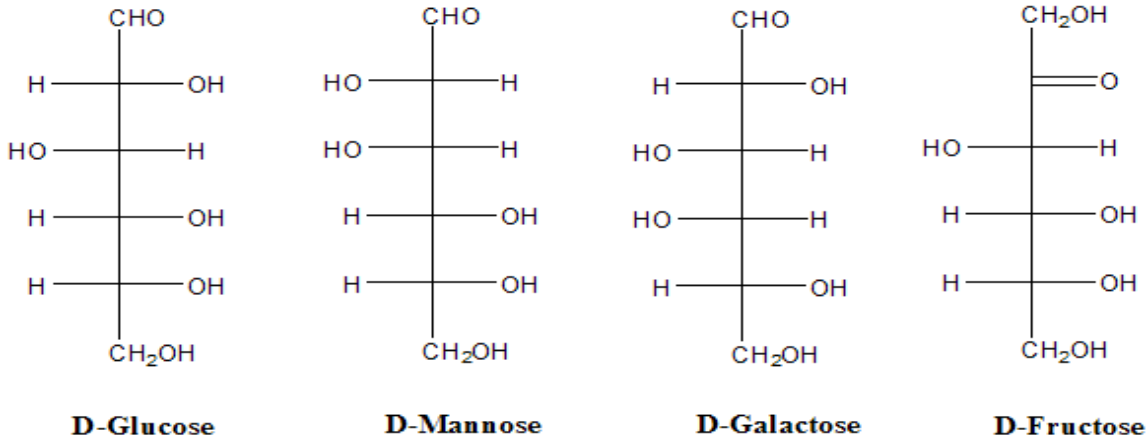


Figure 1.1: Fischer projections of four of the most common hexose sugars found in nature. D-mannose and D-galactose are pivotal building blocks for eukaryotic glycan formation. Image taken from Study blue (2017).

1.2 Glycosylation

In biology, glycosylation refers to the non-template driven process of covalently attaching a series of carbohydrate complexes, or glycans, to a protein, lipid or other organic molecule. The glycans can vary in size from a single carbohydrate to a large oligosaccharide depending on the functionality of the protein or lipid. This process is facilitated by a specific set of enzymes that selectively add and subtract monosaccharides to build the necessary glycan. These enzymes can be split into two distinct groups; these groups are glycosyltransferases and glycosidases. Glycosyltransferases are involved in transferring glycan structures and single sugars to existing glycan units or sites that allow glycosylation (Liarson et al. 2008). These enzymes use a glycosyl donor and join it to a glycosyl acceptor. This process creates a glycosylated acceptor and is explained in figure 1.2. Glycosyl donors, also called nucleotide sugars, are monosaccharides linked to a specific nucleotide.

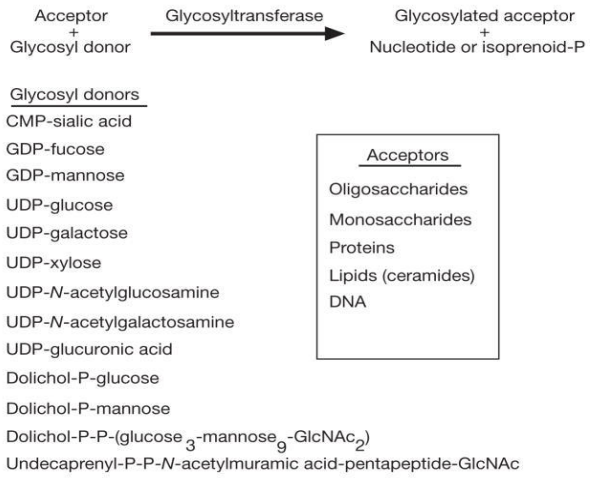


Figure 1.2: Glycoslytransferase activity. Glycosyltransferases use monosaccharides linked to a nucleotide or isoprenoid structure. Image taken from Varki et al. (2009).

Glycosidase enzymes facilitate the subtraction of single monosaccharides or glycan structures from glycosyl acceptors (Boudabbous et al. 2016). Both sets of enzymes are equally important when it comes to formation of the correct glycan structure. Generally in eukaryotes glycans are built up with a higher number of sugar molecules, facilitated by various glycotransferases, than needed. Once these glycans have been built up, glycosidases cut down the glycan and mold it for the desired glycan complex (Bieberich 2014).

It is hypothesised that the majority of our proteins are glycosylated, or have sites that would allow glycosylation. Proteins and cells can undergo many different modifications but studies are showing that glycosylation is one of the most common and also the most diverse form of PTM. In a review by Lommel and Strahl (2009) it is stated that 13 monosaccharides and eight amino acids are involved in at least 41 different types of carbohydrate-protein linkages in eukaryotes. With this information now available it is presumed that if the glycosylation of a protein is incomplete or flawed the resulting protein many have increased or decreased functionality and efficacy (Shental-Bechor and Levy 2008). Alterations in protein glycosylation are a very important area to study

when it comes to disease progression. This area will be explained in further detail in later sections 1.5.1 and 1.5.2.

The intricacies of glycosylation and glycan formation are highlighted by the sheer number of enzymes, sugars and mechanisms involved. To facilitate all of these processes there are various classes of glycosylation. These classes depend on the specific glycosyl acceptor available. These classes are;

- N-linked glycosylation: these glycans attach to the N (nitrogen) group of asparagine or arginine side chains.
- O-linked glycosylation: these glycans attach to the hydroxyl O (oxygen) group which may be present on certain amino acid side chains like serine and threonine. It can also be the classification of glycans binding to the O on a lipid or other organic molecule.
- Phospho-glycans: these bind through a phosphate. Phosphates may be present on a variety of different molecules.
- C-linked glycosylation: this is a rare type of glycosylation where the glycan binds to the carbon side chain on the tryptophan amino acid. This type of glycosylation is rarer due to the fact oxygen and nitrogen are more reactive side groups.
- Glypiation: this is the formation of a glycan link between proteins and lipids. This often allows for the transport between the two molecules.

The last three classes here are usually viewed with less importance due to N- and O-linked glycosylation accounting for the majority of all glycosylation that occurs in eukaryotes. The sugars involved in eukaryotic glycosylation are; mannose, galactose, fucose, *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc) and *N*-acetylnuraminic acid (sialic acid). The addition of Sialic acid to the end of any branch of a complex glycan signals that the glycan has been completed (Lodish et al. 2001). This process is also known as sialic acid capping or just capping. In diseased cells, such as cancer, this capping can occur at the incorrect time or may not occur at all and both of

these can cause significant problems for the protein's intended function. These alterations will be discussed in section 1.5.1.

Phospho-glycans are found in some bacteria as well as certain mammals including humans. Mannose phosphoserine glycans have been found in mice on the cell surface laminin receptor (Yoshida-Moriguchi 2010). C-glycosylation is usually found to be a mannose sugar added to the first tryptophan residue in the sequence W-X-X-W where X is any amino acid (Brenstrum and Brimble 2001).

Glypiation, formation of a glycosylphosphatidylinositol (GPI) anchor, is a very important protein modification. Positioned at the C-terminus of a number of eukaryotic proteins, the GPI anchor facilitates anchoring to the outside of the cell membrane. These GPI anchored proteins have shown to play pivotal roles in a wide variety of biological processes (Paulick and Bertozzi 2008). There are more than 250 known proteins that contain a GPI anchor and are found on the cell surface. Important proteins such as alkaline phosphatase and CD14 are GPI-anchored proteins. The Manα1–4GlcNα1–6*myo*-inositol-1-*P*-lipid is a universal substructure of all GPI anchors in eukaryotes (Varki et al. 2009). This form of "glycosylation" is very important for certain proteins and it helps anchor them to the membrane of the cell. There are relatively low amounts of proteins that contain these types of carbohydrate structures which lead to N- and O-linked glycosylation to be the more important types of glycosylation from an analytical point of view.

1.2.1 N-linked glycosylation

While the three other classes of glycosylation are important in their own right, N- and O-linked glycosylation are the two most prevalent forms of glycosylation (Zu and Ng 2015). O-linked glycosylation will be described in greater detail in section 1.2.2. N-linked glycosylation is the addition of glycans to the N group of asparagine or arginine residues. The targeting of N- glycans is recognised through a specific tri-amino acid sequence, which is asparagine-X-serine/threonine. The amino acid at postion X can be any amino acid with the exception of proline (Varki et al. 2009). In Eukaryotes all N-linked glycosylation occurs in the endoplasmic reticulum (ER) and the golgi apparatus. The first sugar anchored to the asparagine is generally *N*-acetyl-D-glucosamine (GlcNAc). Once this sugar has been successfully attached to the N group of asparagine or arginine, the subsequent glycosylatransferases will facilitate continued glycan synthesis (Aebi 2013). First in the ER a large immature glycan complex is originally added to the specific protein. This allows the protein to fold correctly while some glycosidases begin to selectively reduce the glycan by cleaving individual monosaccharides. Once the protein has folded correctly it is transferred to the golgi apparatus to undergo a number of enzymatic steps to create mature glycans on the protein. Originally in the ER the immature glycan will have a number of mannose residues bound together with various linkages to allow branching of the glycans. The core residues of N-glycans are made up of the same sugar sequence. This sequence is Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asparagine-X-Serine/Threonine.

These N-glycans can be catagorised into three types: the first is oligomannose, this is where the core of the N-glycan consists of only mannose residues with specific α-linkages. The second is a complex glycan; this is where the mannose core has been reduced by glycosidases and other glycotransferases have facilitated the addition of other monosaccharides via specific α- and β- linkages. The final type is a hybrid glycan complex where at least one branch point on the glycan retains its original mannose residues and at least one other branch consists of a number of other monosaccharides

again linked via α- and β- linkages (Varki et al. 2009). There are examples of all three types of these N-glycans in figure 1.3.

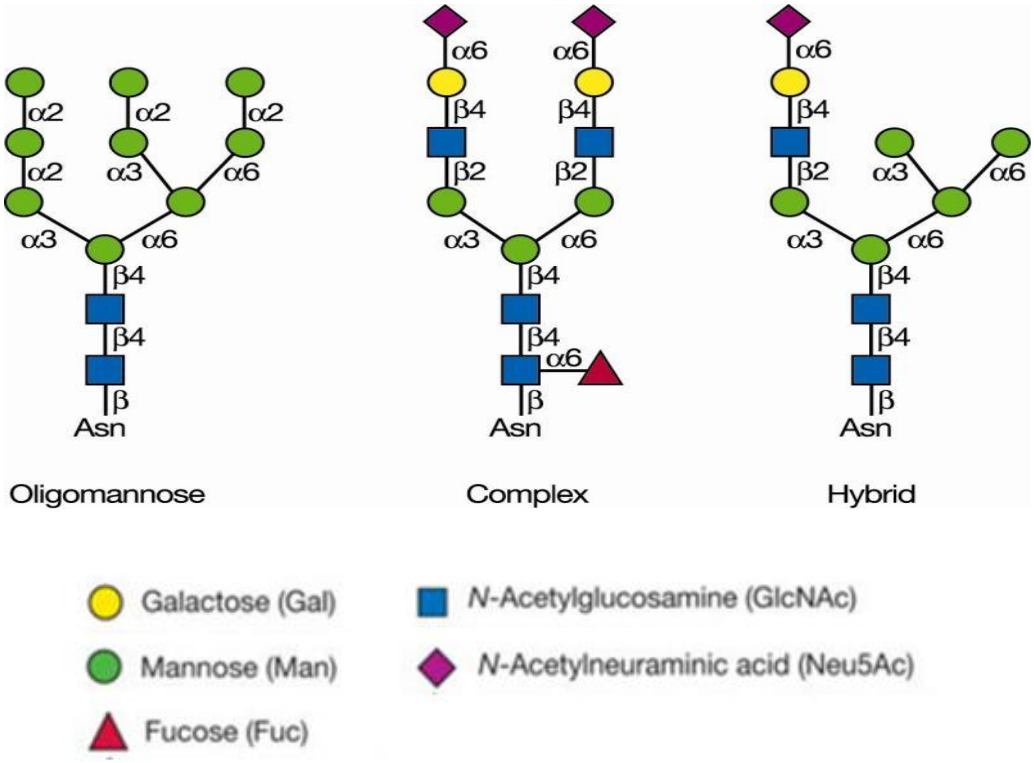


Figure 1.3: Types of N-glycan complexes. This image shows the different types of N-glycans observed in eukaryotes. The specific α- and β- linkages are also shown depending on which sugar molecule has been added at certain positions. Image taken from Varki et al. (2009).

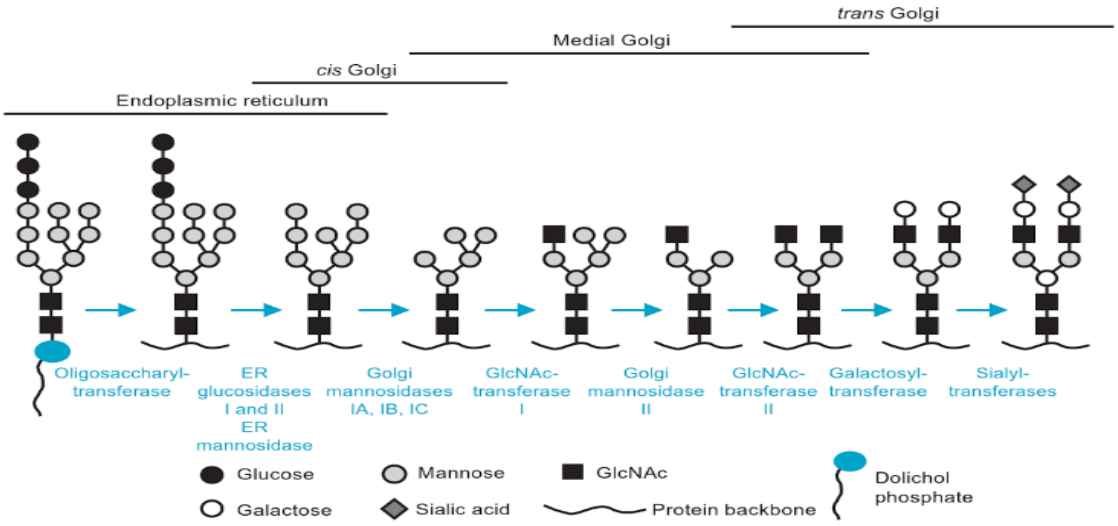


Figure 1.4: N-glycan synthesis from the endoplasmic reticulum to the golgi apparatus. This image depicts the process of creating a mature N-glycan, via glycosyltransferases and glycosidases, in the ER and golgi. Image taken Taylor and Drickamer (2011).

As explained in the previous section 1.2, glycans are anchored to GPI to allow for transport onto an immature protein. This early stage of glycosylation attaches a glycan complex to an unfolded protein. Due to the fact the protein is in an unfolded state, it is assumed that any Asn-X-Ser/Thr site can potentially become glycosylated. There are N-glycan recognition sites that can be buried in folded proteins that do not become glycosylated. Proteins can be modified to add additional glycan sites that would allow for increased N- or O- linked glycosylation. Erythropoietin is an example of a protein that has been modified to have more potential N-glycan sites (Su et al. 2010). As shown in figure 1.4 each individual sugar in added by its own glycosyltransferase enzyme. In mature N-glycans in eukaryotes there are no glucose sugars on the final product. Glucosidases act in the ER to remove the glucose molecules. Mannosidases then act in the golgi apparatus to trim the glycan to allow for the addition of the other sugars. Mature glycans can vary in size anywhere from a few sugars to a >20 sugar complex depending on the level of branching and extension that happens in the golgi apparatus (Grunwald-Gruber et al. 2017).

1.2.2 O-linked Glycosylation

The other significantly important class of glycosylation is O-linked glycosylation. As briefly explained in section 1.2, O-linked glycosylation is the addition of glycan complexes to the hydroxyl group of a serine or threonine residue. There are other amino acids with hydroxyl groups, such as tyrosine, that can also have O-linked glycans, but this is observed less frequently. O-linked glycosylation occurs almost entirely in the golgi apparatus which differs from N-glycosylation (Varki et al. 2009). Early studies of O-linked glycosylation suggested that the majority of glycans began with a GalNAc sugar bound to the hydroxyl group of serine of threonine. Addition of GalNAc to Ser/Thr has been shown to occur only in the golgi. Later in this section it will be highlighted that in the last half a decade the study of O-linked glycans beginning with other monosaccharides has shown this not always to be the case. For instance up to 30% of all brain tissues in mammalian cells are O-mannosylated (Stalnaker et al. 2011). Some of these mannosyltransferases have been found to act in the ER, which changed the long held belief that O-glycosylation only occurred in the golgi and not the ER. O-linked glycosylation has been extensively studied in mucin proteins, as these proteins are secreted by the cell and have been shown to have a very high level of O-linked glycans. There are eight core O-glycans as well as the Thomsen-Nouvelle (Tn) antigen (GalNAc-α-Ser/Thr) and Sialyl-Tn antigen (Sia-α-2,6-GalNAc-α-Ser/Thr). These core structures are shown in table 1.1 and figure 1.5. Mucin type glycoproteins are heavily glycosylated often with short O-glycans (Varki et al. 2009). These proteins were the first observed forms of O-glycosylation and have been studied extensively since their discovery.

Core	
Tn antigen	GalNAcαSer/Thr
Sialyl-Tn antigen	Siaα2-6GalNAcαSer/Thr
Core 1 or T antigen	Galβ1-3GalNAcαSer/Thr
Core 2	GlcNAcβ1-6(Galβ1-3)GalNAcαSer/Thr
Core 3	GlcNAcβ1-3GalNAcαSer/Thr
Core 4	GlcNAcβ1-6(GlcNAcβ1-3)GalNAcαSer/Thr
Core 5	GalNAcα1-3GalNAcαSer/Thr
Core 6	GlcNAcβ1-6GalNAcαSer/Thr
Core 7	GalNAcα1-6GalNAcαSer/Thr
Core 8	Galα1-3GalNAcαSer/Thr

Table 1.1: Structures of O-glycan cores found in mucin type proteins. This image depicts the variety of O-glycans, beginning with GalNAc-α-Ser/Thr, found in mucin proteins secreted by mammalian cells. Table taken from Varki et al. (2009).

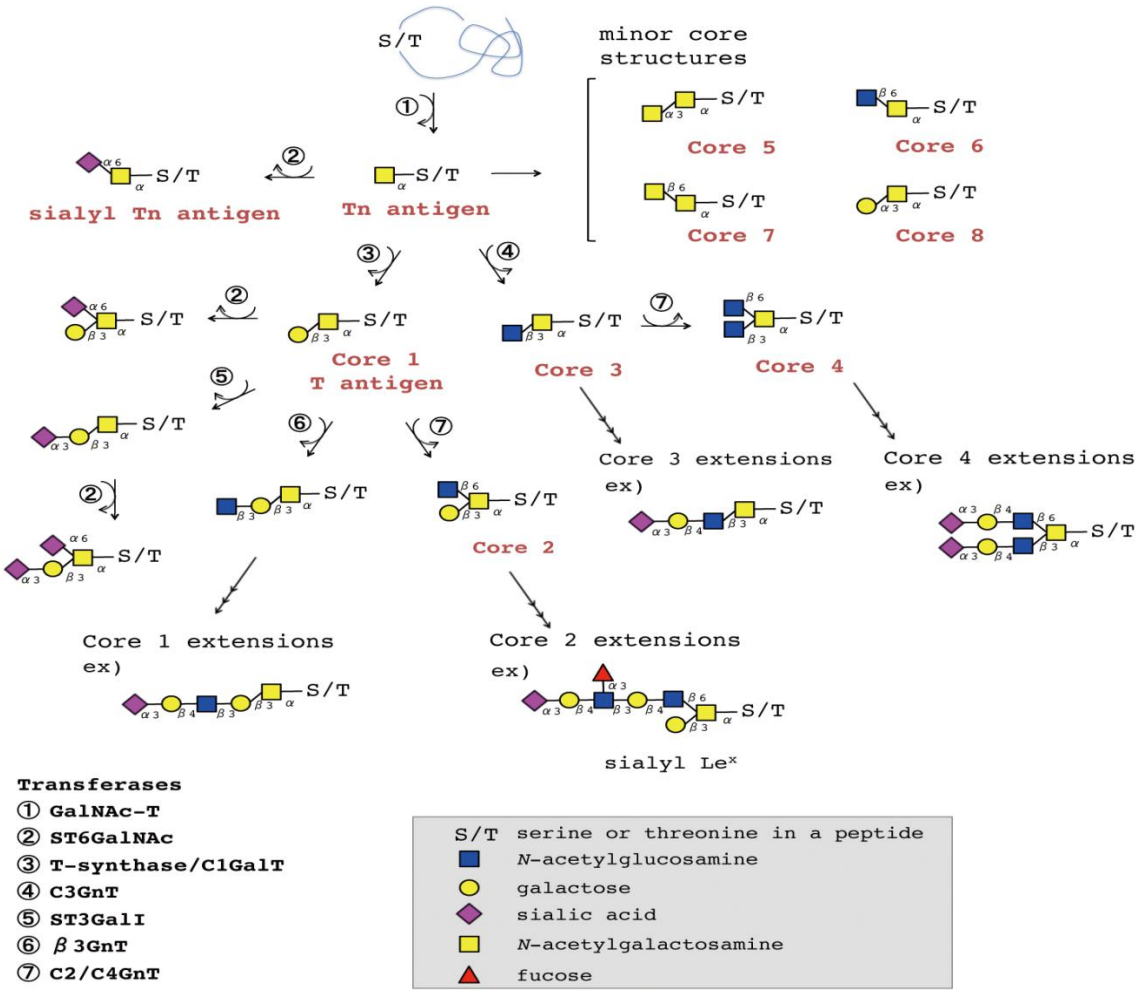


Figure 1.5: Types of O-glycan complexes. This image shows the different types of core O-glycans observed in eukaryotes. The specific α- and β- linkages are also shown depending on which sugar molecule has been added at certain positions. Image taken from Nakayama et al. (2013).

All O-linked glycans can branch in a similar fashion to N-glycans (Figure 1.5). Due to the fact this process occurs primarily in the golgi apparatus, all O-glycans are added to proteins that are already folded. Similar glycosyltransferases are then involved in transferring each individual monosaccharide onto the glycan chain. Differing from N-glycosylation, this process builds up glycans from the base and doesn't use glycosidases

to first trim the immature glycan. Though the study of mucins provided a lot of information about O-glycan sites and O-glycan synthesis, the increased study of glycosylation in the last few decades has shown O-glycosylation to be incredible diverse across different organisms. From prokaryotes to eukaryotes, O-glycosylation has been observed on essentially all amino acids that have a hydroxyl side group (Spiro 2002). In Eukaryotes the primary hydroxyl groups found to have this process occur are shown to be Ser/Thr. Even though a large portion of O-linked glycans begin with a GalNAc, there are more and more studies showing instances of GlcNAc, Mannose, Galactose and even Fucose being the primary sugar at the beginning of the glycan chain (Okajima and Irvine 2002). O-mannosylation has been shown to occur in a wide variety of cell types throughout a number of mammals, including humans. In a recent review carried out by Praissman and Wells (2014) there are a large number of different glycans found to begin with O-mannose. These core structures containing mannose as the primary sugar can be constructed in different ways. For instance a group of core O-mannosylated glycans are all single branched glycans beginning with a GlcNAc monosaccharide bound to mannose via a β1,2-linkage. A very common tetrasaccharide observed in this case is Sia-α-2,3-Gal-β-1,4-GlcNAc-β-1,2-Man (Praissman and Wells 2014). Sialic acid capping of a galactose molecule indicates that there will be no more sugars added to this complex.

Glycosylation with fucose as its primary sugar has been also investigated more thoroughly in the past decade. This is facilitated by two proteins named protein *O*-fucosyltransferase 1 and 2 (POFUT 1 and 2). POFUT 1 was first described in 1998 and the gene was cloned in 2001 and was shown to be pivotal during development of both flies and mice. Later POFUT 2 was identified and characterised in 2006 (Lira-Navarrete et al. 2011). The existence and identification of these two proteins allowed for the study of O-fucosylation to increase. When studying the development of *Drosophila* embryos the Notch protein, and some of its ligands, it was found to be heavily fucosylated. A common tetrasaccharide was found on the Notch protein. This tetrasaccharide is similar to the O-mannosylation described previously; the sequence just replaces the primary mannose with a fucose sugar. The sequence is Sia-α-2,3-Gal-β-1,4-GlcNAc-β-1,3-Fuc

(Okajima and Irvine 2002; Lira-Navarrete et al. 2011). Although O-glycosylation beginning with a primary GalNAc sugar unit has been shown to be the most common version, these other versions of O-linked glycans provide very important roles in a wide variety of proteins.

1.3 Cell Surface Glycosylation

In the previous section 1.2 the process of adding specific sugar molecules to proteins was explained in detail. Once this process is completed and the proteins are packaged and released from the golgi they are translocated to their intended destination, often ending up embedded in the cell membrane. Eukaryotic cell membranes contain a large number of embedded proteins and of these proteins a significant number of them are glycoproteins. The cell membrane's main function is to protect the cell from its surroundings. It's a selectively permeable barrier that can facilitate the movement of ions, organic molecules and other substances in and out of the cell (Alberts et al. 2002). Glycosylated proteins become embedded in the cell membrane and can end up in three different orientations. These orientations are; surface protein with a portion of the protein protruding on the inner membrane, surface protein with a portion of the protein protruding on the outer membrane and the protein protruding on both the inner and outer surfaces, i.e. transport proteins. When a protein is transported and embedded into the cell membrane, its function dictates what orientation it needs to be in. Figure 1.6 shows a eukaryotic cell membrane with numerous proteins infused into the phospholipid bi-layer.

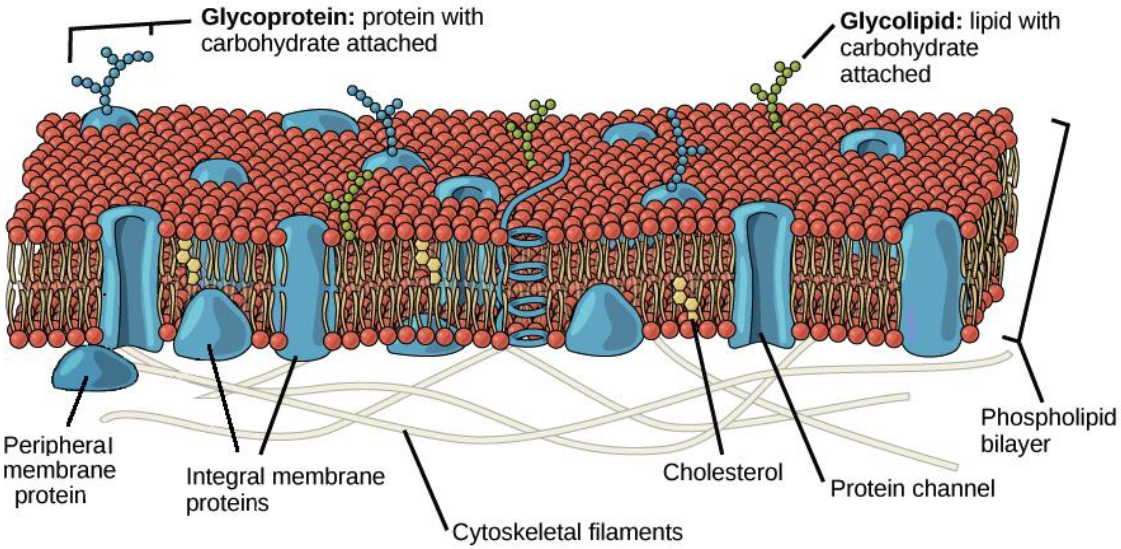


Figure 1.6: Eukaryotic outer cell membrane. Integral membrane proteins are shown to be orientated in a multitude of ways on the inner and outer cell membrane. Image taken from the Khan acadamys online resources (Khan Academy 2016).

In the case of glycoproteins, this outside section of the protein usually contains a number of glycans on its surface. These glycans facilitate a large number of cellular functions. Glycans on the surface of cells are found on receptors, transport proteins and signalling proteins to highlight a few (Christiansen et al. 2014). One of the main functions of cell surface glycosylation in the body is for our own immune system to recognise foreign pathogenic cells while not implicating our own healthy cells. The human immune system recognises the specific orientation of glycans on the surface of its own host cells. When an invading pathogenic cell is then encountered in the body the immune system can recognise it by the unfamiliar glycan pattern on its outer membrane (Turner 2003). This specific immune response is mediated by specific lectins that recognise these unfamiliar glycans and signal the cells to be targeted. This process and the lectins involved will be explained in greater detail in section 1.4. This highlights the importance placed on the cell's ability to correctly glycosylate its surface proteins.

Glycoproteins on the surface can serve a variety of functions. Take the homing cell adhesion molecule (HCAM) protein, or more commonly known as CD44, for example. CD44 is present on the cell surface of the majority of mammalian cell types. In humans the CD44 gene is encoded on chromosome 11 and is a very interesting gene and glycoprotein because of its large number of isotypes, some of which can be seen in figure 1.6. There are 20 exons involved in CD44 and for this reason there are a wide variety of isoforms due to alternative splicing (Goodison et al. 1999). The group of CD44 splice variants play a role in various cellular functions, such as; cell division, migration, adhesion and cell-cell signalling. CD44 is a receptor protein for hyaluronic acid (HA) but can also interact with a number of other ligands. The proteins main role on the surface of cells is to ensure correct cell-cell and cell-matrix adhesion (Ponta et al. 2003). Alternative splicing of the CD44 proteins leads to the addition or deletion of potential glycosylation sites. This glycosylation will impact the proteins ability to carry out its desired function. Some of the variable exons in the CD44 gene encode serine and threonine rich regions which increase the potential amount of O-glycan sites on the protein (Goodison et al. 1999). Due to the proteins nature as an adhesion molecule, CD44 has been increasingly studied in respect to tumour development. The protein has been shown to be upregulated in a number of different cancers (Basakran 2015).

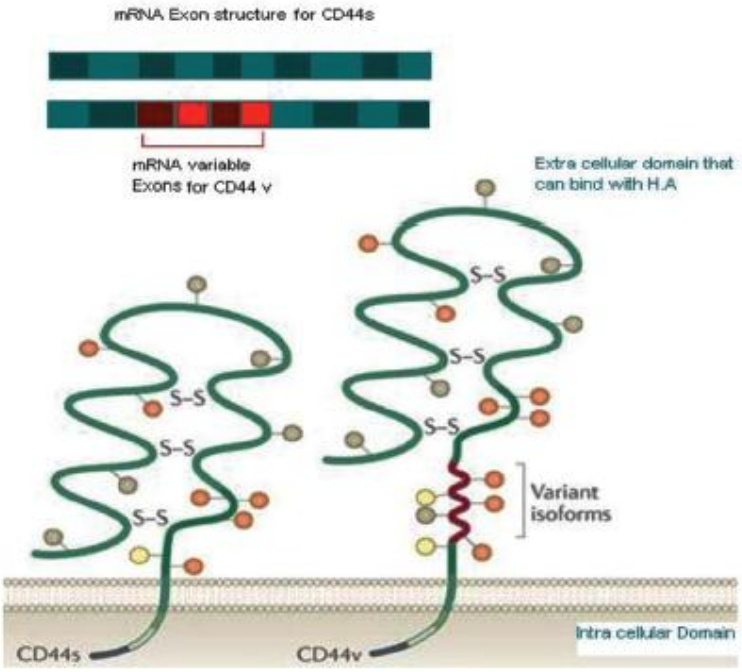


Figure 1.7: CD44 molecular structure of two isoforms of CD44, CD44s and CD44v. Illustration of two of the different CD44 isoforms highlighting the variation of splicing the gene exons. Image taken from review by Basakran (2015).

1.3.1 Current Methods of Analysing Cell Surface Glycans

Researchers have come up with a number of ways to effectively analyse glycans on the surface of cells or proteins. Some of the more important methods used to analyse these oligosaccharide structures are; Mass spectrometry, lectin microarray or lectin blotting, high performance liquid chromatography (HPLC) and the methods used in this project which were live cell lectin probing analysed on fluorescent microscopy and by flow cytometry (Varki et al. 2009). Fluorescent microscopy and flow cytometry will explained in later sections 1.7 and 1.8. There are a variety of ways to analyse glycans and some are more accurate than others, while some are more cost effective and time efficient than others. For this reason the methods are often used in unison and the selection of which method to use can often come down to specific glycans that need to be identified and characterised.

HPLC is an analytical technique that is used to separate, identify and quantify components in a mixture. It uses the simple principles of traditional liquid chromatography but is pressurised to a significantly higher degree. The method uses a liquid solvent that is mixed with a sample that then is pumped into a pressurised column filled with a solid adsorbent material. Each compound or component of the sample mixture will interact differently with the adsorbent material which affects the flow rate. This change in flow rate is what causes the separation of compounds as they flow out of the column and are analysed. HPLC can monitor glycan samples once they are released from the glycoproteins. There are a number of different types of HPLC; size exclusion chromatography, ion exchange chromatography, reverse phase and normal phase that have been used in glycan analysis. These types of chromatography employ different methods to separate compounds based on the specific characteristics of the compounds in the mixture (Sirard 2012). For glycan analysis a specific strategy must be used in order to correctly identify different glycans that may only differ by linkages and or single sugar changes (Gohlke and Blanchard 2008). Using normal phase HPLC (NP-HPLC) it was possible to obtain robust and reproducible separation of N-linked glycans with a high resolution (Campbell et al. 2008). Firstly the glycans needed to be released

from the specific glycoproteins being tested. This step used peptide:*N*-glycosidase F (PNGaseF) to cleave off the intact N-glycan off the surface of the glycoprotein. These released glycans can then be labelled with a specific fluorophore. One such fluorophore used is 2-aminobenzamide (Liu et al. 2010). The addition of this fluorophore allows detection at a femtomolar level. The process can be seen in figure 1.8, where the multistep process to release, label and ultimately detect specific glycans can be achieved using NP-HPLC. Some of the advantages associated with this method of glycan analysis are: (i) charged and neutral glycans can be analysed in the same mixture; (ii) since its one label per one glycan, the relative amount of individuals glycans can be quantified and (iii) this method can separate structures that have the same composition based on their sequence and the linkages involved (Campbell et al. 2008). The disadvantage of using HPLC is that the specific glycoproteins must first be isolated, and then the glycans must be removed intact and analysed. N-glycans containing a fucosylated precursor GlcNAc cannot be cleaved by PNGase F which is another drawback (Tarentino et al. 1989). For these reasons HPLC may often be coupled with other methods of glycan analysis.

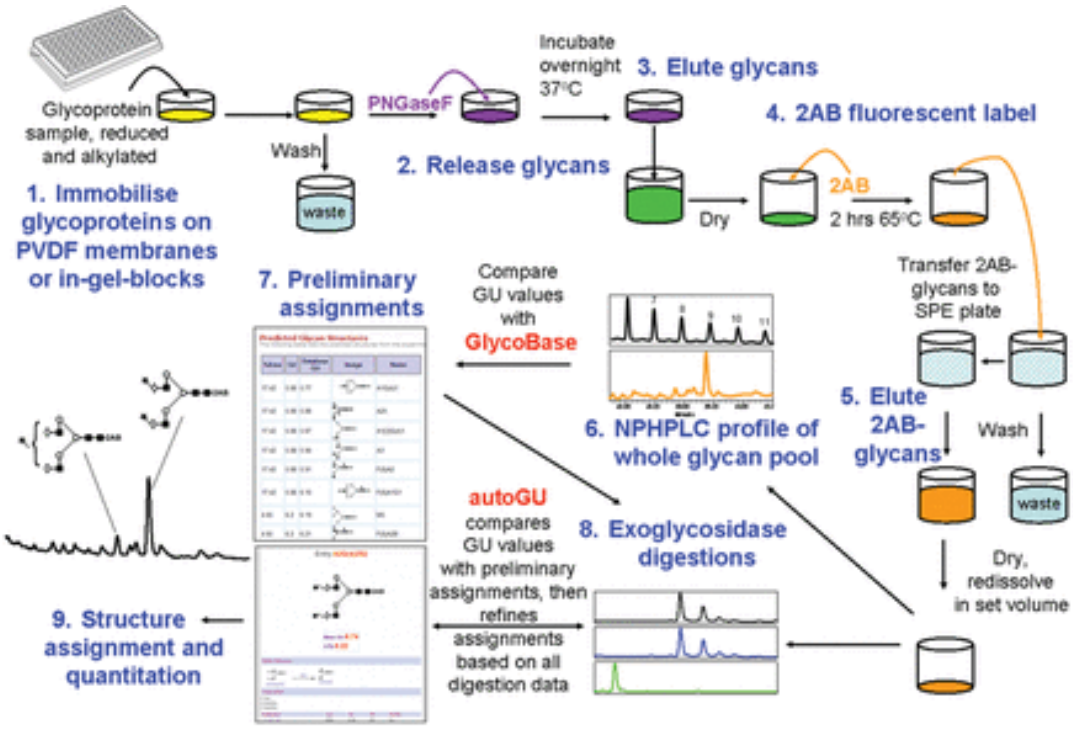


Figure 1.8: Overview of experimental strategy for glycan release, labelling and sequencing of glycans using HPLC. This diagram shows an effective way to identify, characterise and quantify glycans using HPLC. Image taken from Liu et al. (2010).

Lectin microarrays and lectin blotting are methods used to determine the specific glycans on the surface of cells using proteins that specifically bind to carbohydrate molecules. The next section will explain what lectins are in greater detail and talk more about the specifics of certain lectins found both in eukaryotes and prokaryotes. These proteins have unique specificities for determining certain sugar units and their specific linkages. For this reason these proteins can be used in either a microarray format or can be used to blot glycoproteins in the form of western blots.

Microarrays are often described as a lab-on-a-chip platform. The specific capture molecules, in this case different lectins, are immobilised onto the chip in an orientation that allows them to still remain active (Tateno et al. 2007). The large advantage of a microarray is that it can have a large number of different capture molecules on the chip

in duplicate or triplicate to allow for more accurate analysis (Pilobello and Mahal 2007). Specific lectins are chosen to accurately detect every different sugar molecule as well as account for the different branching and linkages that may be present in the glycan structures. For this method the glycoprotein can remain intact and the surface glycans do not need to be cleaved off (Hu and Wong 2009). The experimental overview for setting up a lectin microarray can be seen in figure 1.9. For this method the glycoprotein is labelled with a specific fluorophore and the glycoproteins can then be added to the chip (Hsu et al. 2008). Once the glycoproteins have been captured by the individual lectins, the fluorescence can be monitored through the use of a chip fluorescence reader. Fluorescence or non-fluorescence at a specific lectin site will show the presence or absence of certain carbohydrates or certain carbohydrate linkages (Zhang et al. 2016).

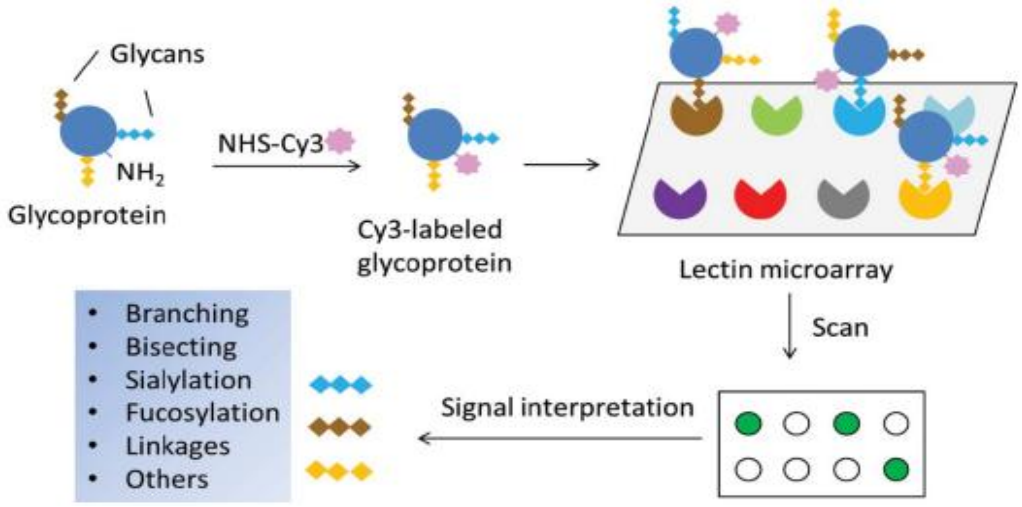


Figure 1.9: Schematic view of lectin microarray. Protein samples are to be labelled with a fluorescent dye (Cy3) and are then applied to the chip that has the lectins immobilised onto it. Image taken from Zhang et al. (2016).

The advantages of using a lectin microarray are: (i) the glycoprotein can be left intact and there is no need to use glycosidases to cleave any glycans; (ii) specific glycoproteins or whole cells can be used on the chip and (iii) a large number of lectins can be used at the same time to account for the large diversity in glycan complexity. Lectin microarrays have proven to be a reliable and accurate tool for detecting carbohydrate complexes and going forward may be one of the more comprehensive ways to identify glycans. The main limitation of lectin microarrays is that the presence or absence of a specific glycan is often given in a strictly qualitative result. The fluorescence monitored on the chip passes a threshold and this allows for a 'yes or no' answer respectively. Limited quantification of the abundance of specific glycans can be a significant drawback in certain analysis.

Lectin blotting is somewhat similar to a lectin microarray in that it uses lectins to probe glycans still attached to their proteins (Sato 2014). For lectin blotting the standard technique of Western blotting is used but instead of probing with a primary and secondary antibody, a lectin is used as the primary probe and possibly a secondary antibody or, in the case of a biotinylated lectin, a streptavidin linked dye. Western blotting is used to measure the presence and amount of certain proteins in a mixture of proteins run on a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel. The gel is used to separate proteins based on size, the gel is then transferred, or blotted, onto a nitrocellulose membrane and the lectins can be used to probe the proteins now in the membrane (Mahmood and Yang 2012). In the same way as lectin microarrays, lectin blotting can use a wide variety of lectins to distinguish between various glycan patterns. The advantages of lectin blotting are: (i) again glycans do not need to be cleaved off in order to analyse; (ii) a large number of proteins in solution can be tested at the same time and (iii) relative quantification of the specific glycoprotein can be observed (Cao et al. 2013). The disadvantages of this method are that the surface of a cell cannot be specifically tested for; whole cell lysates will contain a wide range of cytosolic glycoproteins. Even using membrane preparation methods, the membrane of the cell may present glycoproteins that have glycans from the inner membrane rather

than the outer membrane. In most cases lectin blotting is used as a precursor to other glycan profiling methods such as HPLC and mass spectrometry.

Mass spectrometry has proven to be a very reliable method for glyco-profiling. Mass spectrometry (MS) is an analytical technique that ionises samples and sorts the ions based on their mass-to-charge ratio (m/z). Samples are essentially bombarded with electrons which ionises the components in the sample mixture and then separates them based on ionic charge by subjecting them to an electric or magnetic field (Schaeffer-Reiss 2008). There are a large range of different MS types such as; time-of-flight (TOF), quadrupole, ion traps, tandem MS as well as many others. A mass spectrometers ability to accurately define each component in a mixture of samples makes it a very powerful identification tool. In respect to glycan analysis, the method to prepare the sample is similar to that of HPLC, where the glycans must be cleaved from the glycoprotein before analysis. The general types of MS used for glycan analysis are Matrix Assisted Laser Desorption/Ionisation (MALDI) –TOF and Electrospray Ionisation (ESI). Both of these are soft ionisation techniques, this means that the ionisation process allows for minimal excess energy to be imparted on the samples and thus causes less fragmentation of the glycans (Han and Costello 2013). Once the glycans can remain intact, the molecular ions can be easily observed. Both of these methods can be used to obtain an overall glycan profile. Using different methods to cleave different N- and O- glycans can help increase the efficiency of accurate glycan analysis. For N-linked glycans the regular PNGase-F method can be used to cleave the glycans, whereas O-linked glycans can be cleaved using reductive elimination. Glycans pose a problem when being characterised on MS. Sialylated glycans can lose their sialic acid caps in the ion source or after the ion extraction from this ion source. This loss in sialic acid proves to be an issue when trying to identify specific mature glycan profiles (Morelle and Michalski 2007). There have been various methods used to stabilise the sialic acid linkage for analysis in MS. The most common way used to stabilise sialic acid residues is to form methyl esters. Some methods use methyl iodide, this eliminates the labile proton, where others have used other methylation methods, such as treating the sample with 4-(4,6-dimethoxy-1,3,5-

triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM). Both of these methods have stabilised the sialic acid on N-glycans for analysis using MALDI-MS (Wheeler et al. 2009).

To date MS proves to be the most comprehensive and accurate way to analyse glycans. Through the use of various different types of MS and sample preparation methods nearly all glycan types can be accurately characterised and to a large degree quantified (Morelle and Michalski 2007). The advantages of the different MS analysis methods for glycan characterisation are: (i) detailed analysis of complex glycan structures; (ii) defined linkage analysis; (iii) largely quantitative as well as qualitative; (iv) there are many different types of MS to choose between for specific analysis. The disadvantages associated with this method are: (i) specified proteins must be identified and the glycan must be cleaved off; (ii) cannot be used for live cell analysis; (iii) operation costs are high and require highly skilled users to prepare the samples, accurately design and carry out the method and perform the analysis on the data set provided. Mass spectrometry is currently the best way of identifying specific glycans on glycoproteins. In later sections, flow cytometry and fluorescent microscopy will be explained and it is our goal to highlight in this introduction and throughout the project that these methods can potentially be used as a tool, instead of or in unison with MS, to accurately glyco-profile the surface of live cells.

1.3.2 Limitation of Antibodies as probes for glycan profiling

Antibodies are a widespread analytical tool used in clinical and research applications. Antibodies simply work by binding to a specific antigen or epitope on particular proteins. Antibodies are designed to bind a unique target with a high level of affinity and specificity. Since the emergence of monoclonal antibodies (mAbs) there has been a major increase in the number of different antibodies available for purchase. While this field has grown a huge amount in the last half a century, antibodies specific for carbohydrates have proven to be more difficult.

Antibodies are designed to bind a specific epitope on a protein or bio-molecule and glycans with similar sugar make ups but differing linkages make them a less than ideal target for antibodies. Antibodies that show binding affinities for glycans often bind core N- and O- linked sugars (Bovin 2013). Even with glycans being difficult targets to bind, there are some commercial antibodies that have been shown to be very effective at binding their sugar target. An example of this is anti-GD2 which is used in treatment of neuroblastomas (Yang and Sondel 2010). GD2 is a disialoganglioside, meaning it contains an added sialic acid onto a galactose residue. One of the main issues with antibodies that have been characterised to bind to glycans is that they often bind mixed epitopes. This leads to a less specific identification (Sterner et al. 2016). Antibodies designed to bind specific carbohydrate complexes can be difficult to create and there are often a large number of glycoproteins on the surface of cells that may have the same, or similar, glycans that the antibody will bind to. Lectins (Section 1.4) are evolutionarily designed to bind to glycans.

1.4 Lectins

In biology, lectin is the general term given to any protein that shows high specificity for binding carbohydrate moieties. The term lectin is derived from the latin word that means "to select". There are studies dating back as far as the 19th century characterising lectins, but in the last 30-40 years studies have vastly increased our knowledge of lectins, what organisms produce lectins and what role lectins play in the biological functions of animals and plants. The first Lectin was isolated by the Russian scientist Peter Stillmark in 1888 from seeds of the castor tree *Ricinus communis*. This lectin was named "ricin" (Gorakshakar and Ghosh 2016). Since the isolation of the first lectin there have been tens of lectins identified, characterised and purified for commercial use. Lectins can be as important as the glycans they target in their biological functions. The unique property of lectins is that each lectin has a specific affinity for a single sugar or sugar complex (Maupin et al. 2012). Due to the fact lectins have such a high specificity and affinity for sugar molecules in a complex glycan, lectins have been identified as good analytical tools for research (Hu and Wong 2009). Lectins binding affinity and specificity are often mediated by the presence of a metal ion (Abhilash et al. 2013; Ahmad et al. 1999). These metal ions are usually calcium (Ca^{2+}) or magnesium (Mg^{2+}) as well as others such as zinc and manganese. Lectin binding is facilitated through a series of hydrogen bonds, hydrophobic interactions and metal ions aid in this process (Sabin et al. 2006). As glycosylation became more and more important as a research topic, lectins too increased in popularity. In section 1.3 the methods currently used to detect glycans have employed the use of lectins in various capacities. Originally lectins were thought to only be located in plants but further research into the field began to reveal lectins in a number of organisms including plants, animals, fungi and bacteria (Fujihashi et al. 2003; Taskanen et al. 2001). The increase in the amount of different lectins available for research has helped to continually progress the area of glycobiology and glyco-profiling. Lectins are even being utilised in industrial processes, such as lectin chromatography, to screen bio-pharmaceuticals for the specific glycoforms needed (O' Connor et al. 2017).

1.4.1 Eukaryotic Lectins

In Section 1.2 it has been highlighted that glycans in eukaryotes can greatly vary in complexity and structure. The cells ability to glycosylate proteins in such a multitude of ways allows for a need for highly selective lectin proteins to bind to every possible glycan they encounter. In eukaryotic glycosylation there are six main sugars, described in Section 1.2, that make up a wide variety of glycans through branching and various α- and β- linkages. The first lectins identified and characterised were all from plant sources, leading to a common misconception that lectins were solely confined to plant cells. Some lectins have been shown to be insecticidal when congregated in large numbers. This process helps to protect the plant cells from insects that could potentially destroy the cell. An example of this is Wheat Germ Agglutinin (WGA) which binds β-1,4 linked polyGlcNAc. This lectin is ingested and is poisonous to the insect (Schaller 2008). Due to the protective factor of WGA it is often found in very large quantities in the plants that express the protein. Other plants have lectins that show a similar protective function as a lot of plant lectins will bind the surface carbohydrates of invading pathogens. Some other eukaryotic lectins do not always have an obvious function like WGA. There are a large range of lectins that have been identified, characterised and purified from eukaryotic plant sources. A lot of these lectins have proven to be important research tools, as highlighted in lectin microarrays and lectin blotting. Specific lectins can be used when trying to identify unknown glycoproteins. To test for unknown glycans on a glycoprotein, lectins that bind common glycan complexes can be used, e.g. concanavalin A (Con A). Con A is derived from the jack-bean *Canavalia ensiformis* (Reeke Jr. et al. 1975). It binds to terminal non-reducing α-D-mannosyl and α-D-glucosyl groups (Kim and Park 2001). Con A exhibits a high specificity for the core mannose structures found in all N-glycans. The majority of glycoproteins contain complex N-glycans, while testing for the presence of N-glycans Con A has proven to be an important lectin for identifying this. Table 1.2 shows the variability between different lectins in binding specificity as well as the necessity of a metal ion, size and number of subunits in the protein.

Lectins and their binding specificities

Lectin	Source	Subunits	Mol. weight (kDa)	Glycan specificity	Eluting sugar/analog	Compatible metal ions
ConA	*Canavalia ensiformis*	Four identical subunits	104	α-Mannose, glucose	200 mM mixture of α-methylmannoside and α-methylglucoside	Ca^{2+}, Mn^{2+}
ECL	*Erythrina cristagalli*	Two different subunits	54	Galactose, *N*-acetylgalactosamine, lactose	200 mM lactose	Ca^{2+}, Mn^{2+}, Zn^{2+}
AAL	*Aleuria aurantia*	Two identical subunits	72	α-1,6-Fucose, fucose-linked α-1,3 to *N*-acetyllactosamine	100 mM L-fucose	None
Jacalin	*Artocarpus heterophyllus* (Jackfruit)	Four subunits	66	T-antigen, sialyl-T-antigen, galactose	800 mM galactose, 100 mM melibiose	None
WGA	*Triticum vulgaris* (Wheat)	Two subunits	36	*N*-acetylglucosamine	500 mM GlcNAc with salt and/or acid	Ca^{2+}
LCA	*Lens culinaris*	Four subunits	50	α-Mannose, glucose	200 mM mixture of α-methylmannoside and α-methylglucoside	Ca^{2+}, Mn^{2+}
PNA	*Arachis hypogaea* (peanuts)	Four subunits	110	Galactose (β3 linked to GalNAc)	200 mM galactose	Ca^{2+}, Mg^{2+}
Mal II	*Maackia amurensis*	Two subunits	130	α-2,3-NeuNAc (sialic acid)	Human glycophorin (glycoprotein)	None
LecB(PA-IIL)	*Pseudomonas aeruginosa*	Four Subunits	13	Fucose, mannose Fucose>mannose	200 mM L-fucose, 200 mM D-mannose	Ca^{2+}
BC2LA	*Burkholderia cenocepacia*	Two subunits	14	Mannose	200 mM D-mannose	Ca^{2+}

Table 1.2: Lectins from different organisms. Lectins from plants, fungi and bacteria are highlighted in this table. Along with the lectins name and source this table shows the molecular weight, glycan specificity, eluting/inhibitory sugar and the necessity of metal ions for the lectin. Table obtained from O' Connor et al. (2017).

Other plant lectins have been utilised to distinguish between blood groups in humans. There are four red blood cell types that are associated with humans, these are: A, B, AB and O blood types. Lectins can be used to discern exactly which blood group a person has. A lectin from *Vicia villosa*, named VVA, has proven to be an excellent anti-A blood type lectin. A lectin from *Griffonia simplicifolia*, named GSL I, has proven to be an excellent anti-B blood type lectin (Khan et al. 2002). These examples highlight the importance of certain plant lectins and show how applicable they have been and may prove to be in past and future research.

There is a lectin recognition pathway in the immune system of humans. This mannose binding lectin (MBL) is well studied to this point and proves to be an immensely important molecule in pathogen recognition and immune response. This mannose binding lectin binds to carbohydrates on the surface of bacteria and initiates innate immunity via the complement pathway (Wallis 2007). The understanding of the role MBL in immune response has grown rapidly over the past few decades and it has been recognised to have a role in a number of immune processes. As well as complement activation it has also been found to modulate inflammation, recognise alterations on self-structures and has a role in apoptotic cell clearance (Dommett et al. 2006). Phagocytes have MBL-receptors on their outer membrane surface. The MBL binds to either an apoptotic cell or bacterium and remains bound until it can recruit a phagocyte. Each MBL has a MBL-associated serine protease (MASP) region that the MBL receptor binds to (Kravitz and Shoenfeld 2006). In figure 1.10 the process of MBL mediated phagocytosis is highlighted.

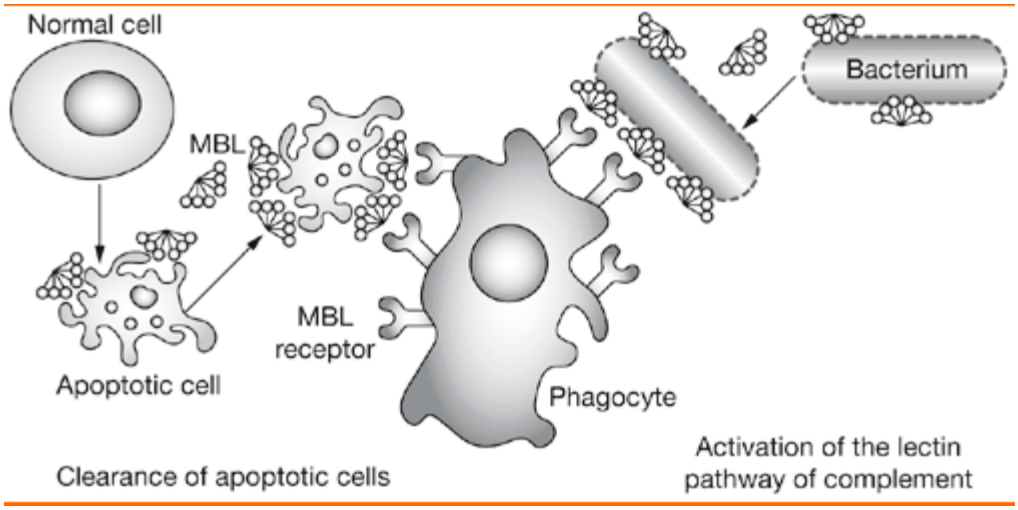


Figure 1.10: Diagram of mannose binding lectin attaching to apoptotic and bacterial cells to initiate phagocytosis. MBL receptors on phagocytes bind to the MASP region of MBL and activate phagocytosis. Image taken from Kravitz and Shoenfeld (2006).

Other eukaryotic lectins of note are *Aleuria aurantia* lectin (AAL) or *Helix pomatia* lectin (HPA). AAL is a lectin purified from the *Aleuria aurantia* fungus and has specificity for fucose linked to GlcNAc or N-acetyllactosamine. HPA is a lectin purified from *Helix pomatia* which is a species of snail. HPA has a dual affinity as it can bind to O-linked GalNAc and O-linked GlcNAc (Ramburuth et al. 2012). HPA also has been shown to bind to Tn antigen, which is an O-linked GalNAc, and is commonly used to screen breast cancer cells. HPA has been used in studies for breast cancer and has indicated that binding of HPA results in a poorer prognosis (Brooks 2000). HPA studies prove lectins potential in determining glycosylation differences on cancer cells and this project will also try and utilise lectins as a probe to distinguish between cancer cell lines. All the lectins mentioned throughout this section are commercially available, and have been proven to be reliable research tools (Hagerbaumer et al. 2015).

1.4.2 Prokaryotic Lectins

Early on in the study of lectins, it was believed that lectins were specifically confined to eukaryotes. Lectins are often large and complex proteins with multiple subunits, some of which need post translational modifications (PTM) such as glycosylation and phosphorylation (Colley and Baenziger 1987). For these PTMs, especially glycosylation, prokaryotic cells have been proven to have less complex modifications after and during protein synthesis. Unlike the other eukaryotic lectins, which mainly need to be expressed in their own host cell, prokaryotic lectins can be sequenced and transformed into a stable bacterial host cell and mass produced much quicker and cheaper. The emergence of various prokaryotic lectins has been important for the lectin and glyco-analysis industry. In prokaryotes, the lectins produced are often smaller proteins with a single binding domain. This is different to a lot of eukaryotic lectins that often have multiple binding sites (Kravitz and Shoenfeld 2006). The mannose binding lectin mentioned in the Section 1.4 has multiple binding sites for both fucose and mannose whereas a prokaryotic lectin, such as PA-IIL which, as a monomeric protein, has one binding site. These smaller proteins, with single binding sites and little to no PTMs have an advantage over eukaryotic lectins for these reasons.

There have been a number of prokaryotic lectins discovered to date. Some pathogenic bacteria use lectins on their outer membrane surface to bind to the sugars on eukaryotic cells. *Pseudomonas aeruginosa* is an opportunistic pathogen that has been studied with respect to chronic airway and lung infections. This *Pseudomonas* strain synthesises two lectins PA-IL and PA-IIL, now named LecA and LecB (Funken et al. 2012). A lectin encoded in *Escherichia coli* called GafD was discovered in diarrhea-associated *E. coli*. This lectin was overexpressed in *E. coli* strains that were associated with pathogenicity causing diarrhea and other illnesses (Tanskanen et al. 2001). These are the lectins that have been discovered in different pathogenic bacteria. The function of these lectins is to bind the complex glycans on the surface of mammalian cells. In table 1.3 these bacterial lectins, as well as others, are detailed and their specificity given.

Lectin name (if applicable)	Source	Specificity
Fimbriae	Actinomyces naeslundii	Galβ1–3GalNAcβ-
FHA	Bordetella pertussis	Galα1–4GlcNac
	Borrelia burgdorferi	Heparan sulfate
	Campylobacter jejuni	Fucα1–2Galβ1–4GlcNAcβ-
P fimbriae	Escherichia coli	Galα1–4Galβ
S fimbriae	Escherichia coli	Gangliosides GM3 and GM2
Type-1 fimbriae	Escherichia coli	Manα1–3(Manα6)Man
K99 fimbriae	Escherichia coli	Neu5Gcα2–3Galβ1–4GlcNAc
GafD	Escherichia coli	β-GlcNAc
	Haemophilus influenza	Neu5Acα2–3Galβ1–4GlcNAc
BabA	Helicobacter pylori	Lewis B antigen
SabA	Helicobacter pylori	Neu5Acα2–3Galβ1–4GlcNAc
	Mycobacterium tuberculosis	Neu5Acα2–3Galβ1–4GlcNAc
	Neisseria gonorrhoeae	Neu5Acα2–3Galβ1– 4GlcNAc
	Propionobacterium	Lactosylceromide
PA-IL (LecA)	Pseudomonas aeruginosa	α-galactose
PA-IIL (LecB)	Pseudomonas aeruginosa	Fucose/mannose
RS-IIL	Ralstonia solanacearum	Mannose/fucose
	Staphylococcus aureus	Heparan sulfate
	Streptococcus pneumoniae	GlcNAcβ1–3Gal

Table 1.3: Recombinant prokaryotic bacterial lectin panel. This table shows the bacterial lectins produced certain bacteria as well as giving the glycan binding specificity. Table generated during this work.

Companies that supply eukaryotic lectins		
EY Laboratories	Vector Laboratories	Sigma-Aldrich
Sino Biological	TCS Biosciences	Thermo Fischer Scientific

Table 1.4: Current companies that sell a multitude of eukaryotic lectins commercially. This table gives the names of six companies that sell commercial eukaryotic lectins for analytical use. Table generated during this work.

1.4.2.1 Prokaryotic Lectins, PA-IL and PA-IIL (LecA and LecB)

The opportunistic pathogen *Pseudomonas aeruginosa* is involved in acute infections as well as chronic infections, especially in cystic fibrosis patients that already have significant issues with other pathogens. This bacterium expresses two lectins, LecA and LecB. These two lectins play a pivotal role in the bacteria's pathogenicity (Chemani et al. 2009). These proteins adhere to the glycoproteins on the outer cell membrane of the host cell and initiate infection. LecA and LecB are found in high levels in the cytoplasm of the bacteria cells as well as large quantities on the outer surface of the bacterial cell membrane (Funken et al. 2012). LecA and LecB have been shown to be virulence factors for *Pseudomonas aeruginosa* and aid the bacteria in producing a biofilm. LecB deficient mutants of this *Pseudomonas* strain have shown an inability to effectively create a biofilm compared to the wild type bacteria (Tielker et al. 2005). These proteins have different binding specificities that have since been identified and the lectins have been characterised. Both LecA and LecB can be produced recombinantly and used as lectin probes for other glyco-profiling procedures. The ability to express these proteins in a bacterial host cell is an advantage of prokaryotic lectins over most eukaryotic lectins.

LecA is a galactose binding lectin. Through various characterization experiments, LecA has been shown to bind to α-galactose, specifically with the linkages α-1-2 and α-1-3. The 3D structure of LecA is shown in figure 1.11. LecA was screened against bovine serum albumin (BSA) with α- and β- linked galactose containing different position linkages, i.e. 1-2, 1-3, 1-4 and 1-6. This characterization helped to promote this protein as a potential research tool in glyco-profiling (Keogh et al. 2014). This has been backed up by other research that shows specificity for terminal α- galactose of different linkages. Some N-glycans have terminal α- galactose and these N-glycans can be found on the surface of eukaryotic mammalian cells. LecA has been compared to other plant

lectins that bind specifically to galactose residues and it has been shown to bind with similar affinity and specificity as those other lectins (Kirkeby et al. 2006).

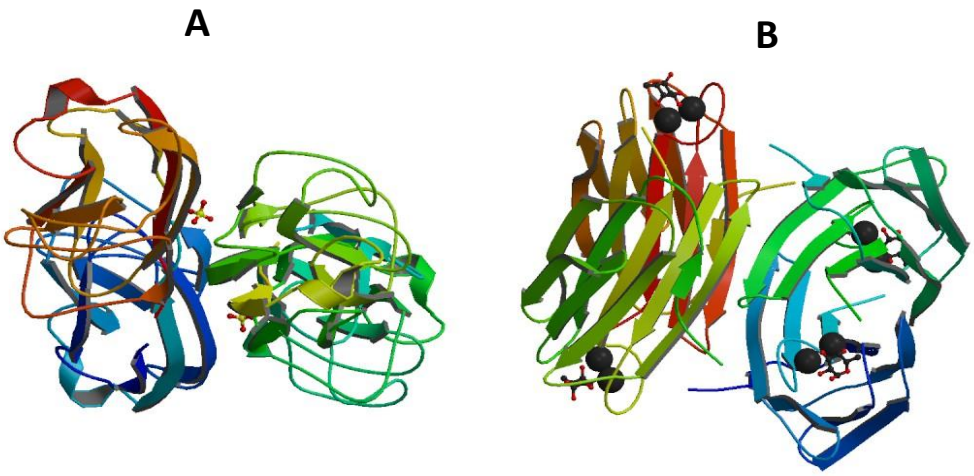


Figure 1.11: 3D structure of prokaryotic lectins LecA and LecB. A) 3D structure of LecA, B) 3D structure of LecB. These images were created using the 3D protein modelling software PyMol.

LecB is the main prokaryotic lectin that has been studied and used throughout the course of this project. LecB is an interesting lectin that has dual sugar specificity, meaning it binds two different sugars with similar affinities. The 3D structure of lecB is shown in figure 1.11. LecB binds L-fucose with a very high affinity and also binds to D-mannose but with a slightly lesser affinity (Hsu et al. 2008; Funken et al. 2012). LecB has a higher proclivity for fucose than it does for mannose but both of these monosaccharides may be used in inhibiting potential binding at certain concentrations. The binding of LecB to any carbohydrate complex or free sugar is mediated by calcium binding to the protein. LecB, as well as many other prokaryotic and eukaryotic lectins, requires a metal ion to facilitate binding. With lecB the metal ion is calcium, without calcium the protein cannot bind to its target (Sabin et al. 2006). The interaction between the protein and carbohydrate is established through coordination by calcium ions and hydrogen bonds

(Loris et al. 2003). Other lectins often establish binding through hydrophobic interactions, but lecB does not rely on these interactions to the same level. This makes lecB a unique and interesting protein while also highlighting the importance of metal ions in respect to lectin binding.

Both lecA and lecB have been studied for their role in pathogenicity but in more recent studies these proteins have been identified as unique lectins for research and industry purposes. In the next section the production of recombinant lectin probes will be explored and both of these proteins have proven to be very good candidates for recombinant production.

1.4.3 Recombinant protein production in bacterial host cells

One of the main advantages of prokaryotic lectins over eukaryotic lectins is that the proteins can be expressed in a bacterial host cell to achieve high levels of expression. Most eukaryotic lectins require eukaryotic PTMs which mean they cannot be produced from a bacterial host cell. Expression systems for recombinant eukaryotic proteins are often transformed into a eukaryotic host cell, such as Chinese hamster ovary (CHO) cells, so that all the complex PTMs are carried out (Wurm 2004). The use of mammalian cells adds a lot of extra costs and it's a more complicated purification process due to the growth conditions of mammalian cells. Prokaryotic lectins, such as lecA and lecB, can be easily expressed in a bacterium, such as *E. coli*, and high yields of protein can be easily purified. In the majority of cases *E. coli* is used as the organism for expression. There are other organisms that can be used, ranging from bacteria to unicellular algae. *E. coli* has a number of advantages over other microbes when it comes to using it as an expression system. With optimal environmental conditions, the doubling time of this bacterium can be as little as 20 minutes (Sezonov et al. 2007). A large cell mass can be grown in a relatively short time, i.e. 6-24 hours. Other advantages for using *E. coli* are; the media needed for cellular growth is readily available and inexpensive and *E. coli* is a host that can have DNA plasmids easily transformed into the bacterium (Rosano and Ceccarelli 2014).

Once the host organism is chosen there are other factors that need to be addressed for the protein of interest that is to be expressed. As mentioned *E. coli* is easily transformable so an appropriate plasmid vector must be chosen to carry the gene that expresses the desired protein, i.e. the lecA or lecB gene in the case of a prokaryotic lectin. Plasmid vectors are self-replicating, extra-chromosomal DNA loops found inside certain bacteria (Lodish and Zipursky 2001). There is a large library of vectors that have various characteristics that may be desired. Plasmids have selectable traits that may be necessary for selection purposes. Firstly, antibiotic resistance allows for selective growth of the bacteria. *E. coli* does not have natural antibiotic resistance, but if there is antibiotic resistance on the plasmid transformed into the host then it can be selectively grown on

nutrient agar containing the relevant antibiotics (Bennett 2008). As well as an antibiotic resistance gene, the plasmid must have a multiple cloning site (MCS) that allows for the addition of your specific gene and a selectable promoter region upstream of the protein. The MCS allows for a number of restriction enzymes to be used to open the plasmid and allow DNA to be ligated into the plasmid before transformation into the expression host. The promoter region of a plasmid needs to be upstream of the gene of interest so the gene can be selectively overexpressed to produce protein. By far the most common prokaryotic promoter used is the *lac* promoter or *lac* operon. The *lac* operon is the region of DNA in bacteria that facilitates the breakdown of lactose in the absence of glucose. *E.coli* contains the lac operon in its bacterial DNA. The inserted plasmid contains two lac operator regions where the lac repressor protein can bind. By using a lactose analogue, such as isopropyl β-D-1-thiogalactopyranoside (IPTG), the repressor can be constantly bound and this allows for unimpeded translation of the genes in the lac operon (Rosano and Ceccarelli 2014). This promoter region can be used in specific plasmids to ensure over expression of the gene added to the MCS. The last important trait that can be either added to the gene specifically or found on the plasmid is a 'tag' region to allow for affinity purification of the protein produced. In plasmids such as pQE30 and pQE60, there is a poly-histidine DNA region either upstream or downstream of the MCS to allow for an *N*- or *C*- terminally linked 'His' tag (Huang et al. 2006).

In this project the plasmids selected for lecA and lecB expression were pQE30 and pQE60 that covered all the desirable traits needed for protein expression. These plasmids contained ampicillin resistance, a lac operon promoter region, MCS and tag on the *N*- or *C*- terminus of the protein. In figure 1.12 the gene map for pQE30 shows the desirable features contained within the plasmid.

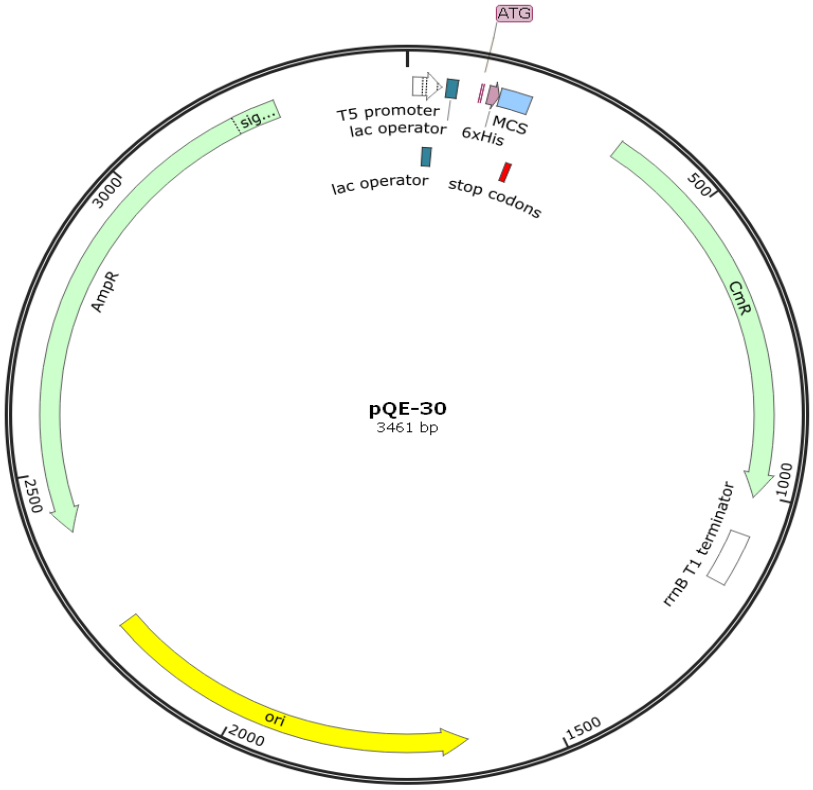


Figure 1.12: pQE30 plasmid vector map. The pQE30 vector contains a number of desirable features for recombinant protein production such as the multiple cloning site (MCS), lac operator, T5 promoter, ampicillin resistance gene and the poly-histidine tag. Map generated using snapgene (Section 2.30).

Proteins containing a poly-histidine tag can be easily purified in what can be described as a 'single step' purification by using immobilised metal affinity chromatography (IMAC). In this method metal ions such as copper, zinc and nickel are immobilised on activated resin and a solution containing a protein mixture is added to the resin. The poly-histidine tag has a high affinity to the metal ion and is held in complex with the metal ions. All other proteins without a poly-histidine tag will not interact, or will have very weak interactions with the metal ion attached to the activated resin. This version of chromatography is usually carried out in a specific column using resin bound with the metal ion. The protein samples are run over the resin and the resulting flow through contains the proteins that didn't bind to the resin. The interaction between the metal ion

and the histidine is the protonated imidazole side group on histidine (Bornhorst and Falke 2000). The use of free imidazole can then selectively elute the protein via competitive inhibition. Certain proteins may contain a naturally high level of histidines and may bind weakly to the resin, so low level imidazole washes are usually carried out before elution of the protein. The relative ease of purification of the tagged protein is a major advantage when using a prokaryotic expression organism.

A significant amount of research has been carried out in respect to protein expression and purification. Through using *E. coli* as an expression host, a pQE30/60 plasmid as the expression vector and poly-histidine tag as the specific tag to aid in purification, any protein can be easily and inexpensively overexpressed and purified. Employing this method both lecA and lecB could be successfully expressed and purified.

1.4.4 Advantages of recombinant prokaryotic lectins over commercial eukaryotic lectins

Recombinant protein production has revolutionised researchers and industry's ability to mass produce certain proteins quickly, efficiently and with very little relative cost. For research purposes the use of recombinantly produced lectins over eukaryotic commercially purchased lectins is advantageous. Lectins purified from plants that are bought commercially have some significant disadvantages. The most important disadvantages stem from a low yield of lectin from its original source organism as well as batch-to-batch variation of wildtype lectins. Eukaryotic lectins can come from a number of organisms, such as plants, animals and fungi. These organisms that produce lectins, express the lectins at low levels to facilitate their biological function. Lectins obtained from seeds often found are in higher levels than those found in plants and animals (Lam and Ng 2011). Purifying lectins from their host organism can be difficult and require a multi-step purification that gives lower yields of protein. This in turn drives the cost for such lectins up and therefore makes them less viable as a large scale diagnostic tool. The other main disadvantage that eukaryotic lectins have over prokaryotic lectins is the fact that they often vary from batch-to-batch (Oliveira et al. 2014). Lectins isolated from their natural source are often a heterogeneous mixture of lectin isoforms that each serves a specific biological function (Han et al. 2015). When using lectins as bio-diagnostic tools, this heterogeneity can cause varying results and this can result in significant issues. Also, as mentioned previously, the majority of eukaryotic lectins need eukaryotic post translational modifications, which means they cannot be produced recombinantly in a bacterium.

The disadvantages for large scale production of these plant and animal lectins show the advantages of producing recombinant lectins from a microbial host organism. Batch-to-batch variation has shown to not be an issue when producing recombinant prokaryotic lectins, such as lecA and lecB. These proteins are routinely expressed, purified and characterised and show the same level of binding specificity and affinity with each batch

(Keogh et al. 2014). The ability to produce non-varying batches of lectins increases their potential as biomedical and bioanalytical tools. In the previous section the selectable traits that are desired for recombinant protein production were discussed. Through the use of a selectable promoter and affinity purification tag, proteins can be programmed to overexpress inside the bacterial host strain and then once the cells have been lysed the protein may be efficiently purified with a single purification.

Another major advantage of recombinant prokaryotic lectins is their amenability for mutagenesis. Through site directed mutagenesis a library of clones can be created for these lectins to try and improve the lectin. The improvements that can be selectively made using mutagenesis can range from trying to increase the stability or half-life of the protein, to increasing binding affinity or changing binding specificity (Keogh et al. 2014). Subtle changes in the amino acid composition of the protein may lead to vast changes in binding affinity and protein solubility and stability. Employing site directed mutagenesis coupled with targeted 3D modelling of prokaryotic lectins, these proteins can be optimised as bio-analytic tools that would surpass some eukaryotic lectins.

The search for efficient ways to detect glycans attached to glycoproteins, either cytoplasmic/ secreted proteins or embedded outer cell membrane proteins, has utilised a lot of different techniques. One of the drawbacks of using lectins in the past has been the batch-to-batch variation of eukaryotic lectins as well as the low yield produced in their host organism. Identification of lectins from bacteria has lead researchers to believe there are possibly more lectins out there that have various specificities. In the future replacing the use of eukaryotic lectins with prokaryotic lectins that have the same specificities may lead to an increased use in lectins for comprehensive glycan analysis.

1.5 Cancer, a brief overview

Cancer has been and is one of the most well studied diseases to date. Cancer is defined as a group of diseases that involve the abnormal growth of cells with the ability to invade or proliferate into other parts of the body (National Cancer Institue 2017). The first distinction that needs to be mentioned is that not all tumours are cancerous. A cancerous tumour signifies the ability of the tumour to proliferate and spread throughout the body. Benign tumours that do not spread to other organs are often not classed as cancerous. The latest complete figures from the World health organisation (WHO) regarding cancer are for 2015. The five most common cancers between both men and women in 2015 were; Lung, Breast, Colorectal, Prostate and Stomach (World health organisation 2017). In 2012 alone there were approximately 14 million new cases and 8.2 million cancer related deaths. This death toll rose to 8.8 million cancer related deaths in 2015. The majority of deaths from cancer are still observed in the developing world but cancer rates are increasing, not decreasing, in developed countries. The leading causes for this are; tobacco related, alcohol related and dietary. With the advancement of modern medicine it is surprising that the rate of cancer increases year to year. This could be due to better identification processes and awareness, but there are indications that the increase is due to the fact people in developed countries live an unhealthy lifestyle. The fact cancer rates are raising, highlights that more research needs to be done to identify and ultimately treat the more common forms of cancer. During this project the specific type of cancer studied was colorectal cancer. In section 1.5.3 this will be explained in more detail.

There are believed to be six ‘hallmarks’ of cancer. These are; sustaining proliferative signalling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and activating invasion and metastasis (Hanahan and Weinberg 2000). These six characteristics of cancer progression are shown in figure 1.13. Along with these six hallmarks, there is significant evidence to suggest there are two more that could be added to the list, these are; reprogramming of energy metabolism and evading immune destruction (Hanahan and Weinberg 2011). Essentially a cancers

ability to become a sustainable growth inside the body is facilitated by its need to remain alive even though the cells have forgone their primary function.

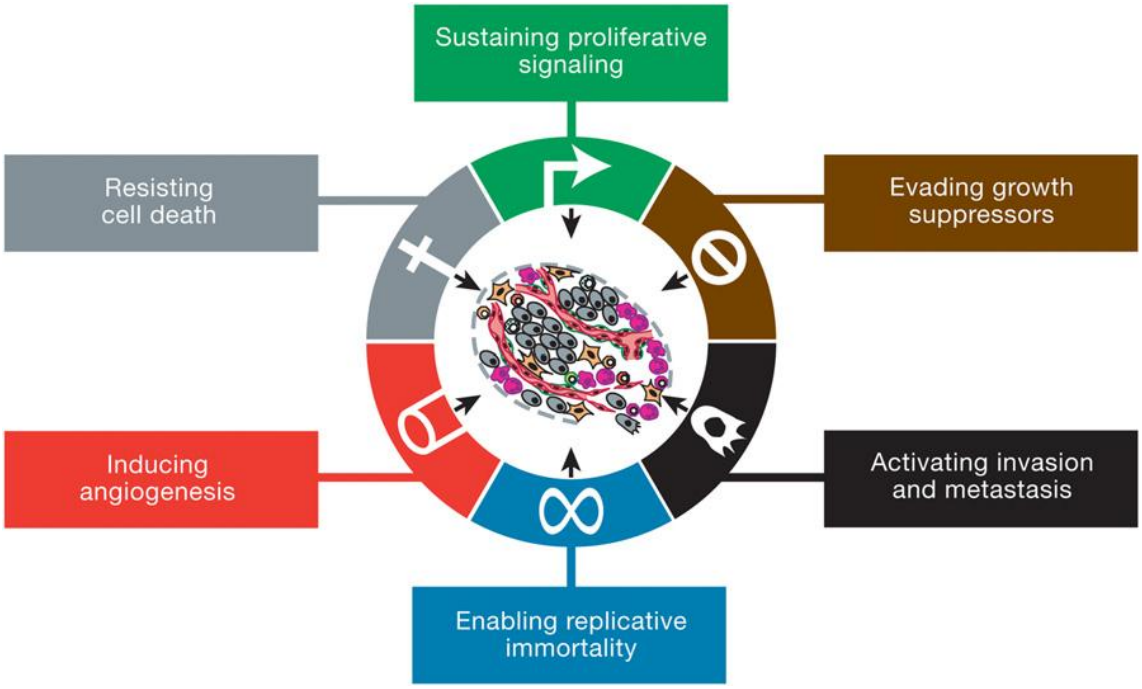


Figure 1.13: Illustration of the six hallmarks associated with cancer. This image shows the original six hallmarks believed to characterise cancer in the early 2000's (Hanahan and Weinberg 2000).

For a cell to become cancerous it must first give up its primary function and try and fit another, self-reliant function. As explained section 1.2, glycosylation plays a big role in many different cell functions from cell adhesion to cell-cell signalling and interactions. This would allow the idea that glycosylation, or altered glycosylation, may prove to be a reliable and important biomarker for cancer progression.

1.5.1 Alterations in glycosylation

Glycosylation is defined as a non-template driven post translational modification. This process is driven by the specific glycosyltransferases and glycosidases mentioned in section 1.2. If these enzymes are either up or down regulated the ensuing glycosylation can be significantly altered and will affect the protein from fulfilling its specific function. In the case of N-glycosylation, alterations to this process may result in incorrect folding of the protein. Before the 1990's there was less understood about glycosylation and alterations in the process were much harder to identify and characterise. Numerous biological functions have been shown to be heavily dependent on correct glycan formation and a complete loss in the cells ability to glycosylate proteins correctly would result in the cell being unable to fulfil its function, or more often than not the cell or organism would not survive (Lyons et al. 2015). Complete loss of function for this reason has proven to be very rare. There is however instances where the glycosylation process is altered due to mutations or illness and the mechanisms are impacted.

In the immune system there has been evidence to show that glycans appear to dictate a variety of aspects of the innate and adaptive immune system. In the innate immune system of humans the first process involved is to recruit immune cells to the site of infection. This recruitment of immune cells is facilitated by the inflammatory response. At this stage of immunity the cells that are first recruited are called leukocytes. Leukocyte rolling is the process of leukocytes connecting to selectins on the surface of endothelial cells close to the infection site (Maverakis et al. 2015). This process is mediated by glycoproteins on the surface of the vascular endothelium. Through specific mutations, the SLC35C1 mutation for example, the glycosylation on the surface of the endothelial cells is altered and the host is impacted. People with this mutation have what is known as leukocyte adhesion deficiency type 2 (Gazit et al. 2010).

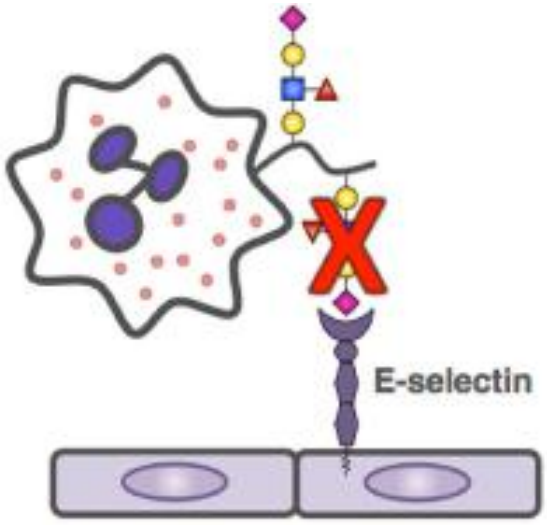


Figure 1.14: Illustration of Leukocyte failing to bind to E-selectin on the surface of vascular epithelial cell. This image shows that the incorrect glycosylation of cell surface glycoproteins on leukocytes can impact their ability to function (Lyons et al. 2015).

Glycans have shown to be pivotal to a number of aspects in adaptive immunity also. For instance with B- and T- cells there are antigen receptors on the surface of these cells that are glycosylated and any alterations in the glycosylation can impact the cells ability to function. CD4 and CD8, which are important co-receptors for T cells, are glycoproteins and there is ample evidence to suggest that these must have the correct glycans on the surface to function correctly (Lyons et al. 2015). Highlighting the significant effects altered glycosylation can have on the immune system's ability to effectively deal with infection, in both innate and adaptive immunity, shows how important correct glycan synthesis is. These instances of alteration in glycan synthesis generally stem from a mutation in a gene expressing a specific glycotransferase or glycosidase. Another way the glycosylation mechanism is impacted is through disease. When a cell becomes diseased it can often change its functionality and because of this it is believed altered glycosylation has a large role to play in how a cell may go about changing its function.

1.5.2 Alterations in glycosylation, specifically in cancerous cells

When a cell becomes cancerous it forgoes its original function. Sections 1.2 and 1.5.1 have explained how important the role of correct glycosylation is for the cells ability to carry out its specific function. These two points together would suggest that in order for cells to change their functionality they would first need to change their surface glycans. In a wide variety of cancers studies have been carried out to identify possible glycan changes on surface proteins as well as proteins secreted by the cell. Secreted proteins such as prostate specific antigen (PSA) and human chorionic gonadotrophin (hCG) are currently used as biomarkers to identify prostate and uterine cancers, respectively. PSA and hCG are both glycoproteins. The current method for identifying prostate cancer is to measure the amount of PSA in a sample of blood taken from a patient (Milford Ward et al. 2001). Prostate cancer is annually the most common newly diagnosed cancer among men (Drake et al. 2015). PSA is a serine-protease that is secreted by epithelial prostate cells. It has been identified as a biomarker due to the fact its secretion levels are elevated in prostate cells that have entered a cancerous state. As one of the first biomarkers identified and easily detectable it garnered wide spread use. PSA screening is still a commonly used identification method to date but in the few decades since it became a routine check, the drawbacks associated with using PSA were identified. Firstly the natural elevation of PSA secreted with increased age. According to the Irish Cancer Society the normal range for PSA in a blood sample raises from 2.5 ng/ml at age 40 to 6.5 ng/ml at age 70. When trying to identify if a person has prostate cancer or not the levels of PSA can vary from 4-100 ng/ml. For this reason there has always been a high rate of false positives in prostate cancer diagnosis. Some elevated levels of PSA can signify benign prostate hyperplasia (BPH) which is a non-cancerous growth. In an attempt to overcome this high rate of false positives there have been many studies carried out to identify possible changes in PSA that may be synonymous with cancer. In a study carried out 15 years ago, through mass spectrometry analysis, there were differences found in the N-glycans present in secreted PSA from normal prostate cells and cancerous cells (Peracaula et al. 2003). Most N-glycans in vertebrates do not contain

GalNAc, though there are numerous exceptions in different animals. N-glycans on PSA commonly contain up to 20% GalNAc but analysis of PSA secreted from cancerous cells can have N-glycans made up of up to 65% GalNAc (Peracaula et al. 2003). Along with differing levels of fucose and sialic acid, the PSA glycoproteins secreted by cancerous prostate cells have distinct carbohydrate differences than PSA secreted from healthy cells. In more recent studies there has been increased emphasis on the level of core fucosylation and sialyation in diagnosing prostate cancer. A recent study from 2016 has identified that PSA secreted from prostate cancer cells has decreased core fucose and increased α-2,3-sialic acid percentages when compared to healthy prostate cells and BPH cells (Llop et al. 2016).

Human chorionic gonadotrophin is another potential glycoprotein biomarker that shows elevated secretion levels in uterine cancer. This hCG protein is commonly used for pregnancy confirmation. Many standard pregnancy lateral flow detectors measure the level of hCG in urine samples of women. Increased levels of hCG are associated with pregnancy but have also been shown to be elevated in certain choriocarcinomas (Stenman et al. 2004). Due to the fact hCG is upregulated in pregnancy and chriocarcinomas are not very common, the level of hCG alone cannot signify the presence of cancer. This is where the analysis of the glycoprotein itself can be used to identify the presence or absence of cancer. When the glycan structures on the hCG proteins were analysed between healthy pregnant women and women who had been diagnosed with chriocarcinoma, there were a number of differences found. The hCG protein is a dimer made up of an α and β subunit, each subunit contains two N-glycans. In healthy pregnant women the N-glycans on the α- subunit were never shown to be fucosylated and often showed minimal to no branching, i.e. the glycans remained mono-antennary. There are different forms of choriocarcinomas, these are; hydatidiform moles and invasive moles. Hydatidiform moles are considered to be benign lesions, whereas invasive moles are considered to be more metastatic with the ability to invade surrounding tissues. The hCG from hydatidiform moles were glyco-analysed and showed the same glycan structure as the hCG secreted from healthy pregnant women.

The interesting changes appeared in the cases of invasive moles. In figure 1.15 there are eight glycan structures shown, labelled A through H. The normal hCG only appears to have glycans E, F and H, whereas the invasive molar cells secrete hCG that has all eight glycan structures (Kobata and Amano 2005). This highlights the upregulation of two glycosyltransferases. One of these glycosyltransferases facilitates additional fucosylation. The other glycosyltransferase forms the GlcNAc-β-1,4-Man-α-1,3 group found in glycan structures A-D in figure 1.15. This linkage does not appear in hCG secreted by healthy pregnant women and therefore can be fully associated with the progression of the disease.

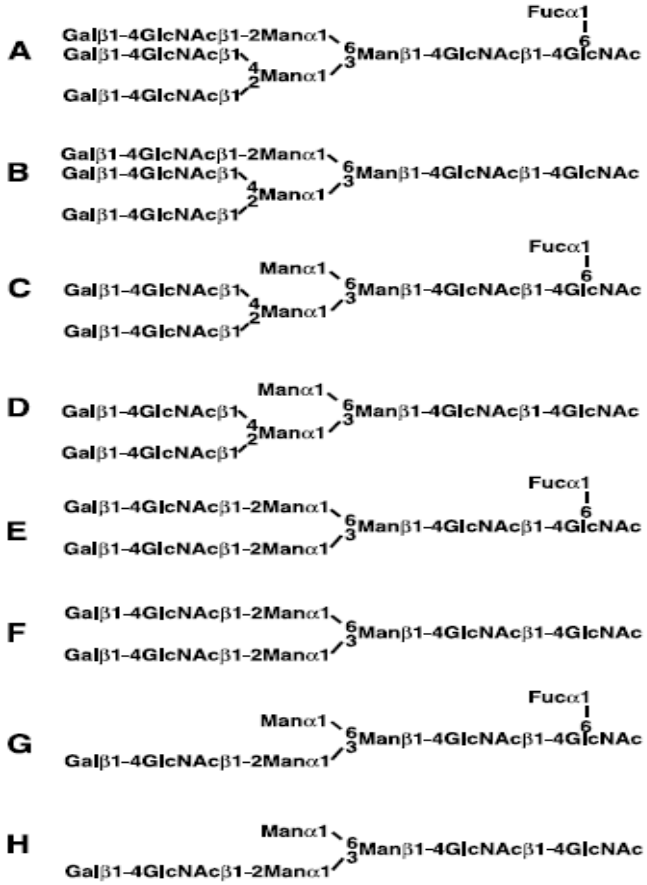


Figure 1.15: Glycans found on hCG protein in healthy and choriocarcinoma cells. This image shows eight glycans found through analysis of N-glycans on the hCG glycoprotein when investigating uterine cancer (Kobata and Amano 2005).

These two examples show how a secreted glycoprotein can be used to analyse the type and progression of certain cancers. Not all cancers will secrete a glycoprotein that can be easily targeted as a biomarker but the fact cancers have been shown to upregulate certain glycosyltransferases gives a mandate for other studies.

As explained in the previous sections, cell surface glycosylation is pivotal to cell functionality. If the glycan structures on secreted proteins can change, then it can be deduced that the glycan structures of the cell membrane glycoproteins may also be altered.

One glycan that is found commonly on the surface of various cancers is the Thomsen-nouvelle (Tn) antigen. The Tn antigen is the most basic form of glycosylation and represents cells that are growing too quickly in a very simple way. The Tn antigen is an O-glycan made up of just a single GalNAc sugar on a serine or threonine (Varki et al. 2009). This is caused by alterations in expression levels of both GalNAc transferases and other transferases that may build up this glycan. There are variations of the Tn antigen, namely the T-antigen and the sialyl Tn or T antigen (Loureiro et al. 2015). Sialyl Tn antigen is a disaccharide glycan which contains a single sialic acid residue attached to the first GalNAc. T-antigen is another disaccharide glycan which contains a single galactose residue attached to the first GalNAc. All of these structures can be seen in figure 1.16. There is the sialylated version of T-antigen also where it has a sialic cap making it a tri-saccharide glycan.

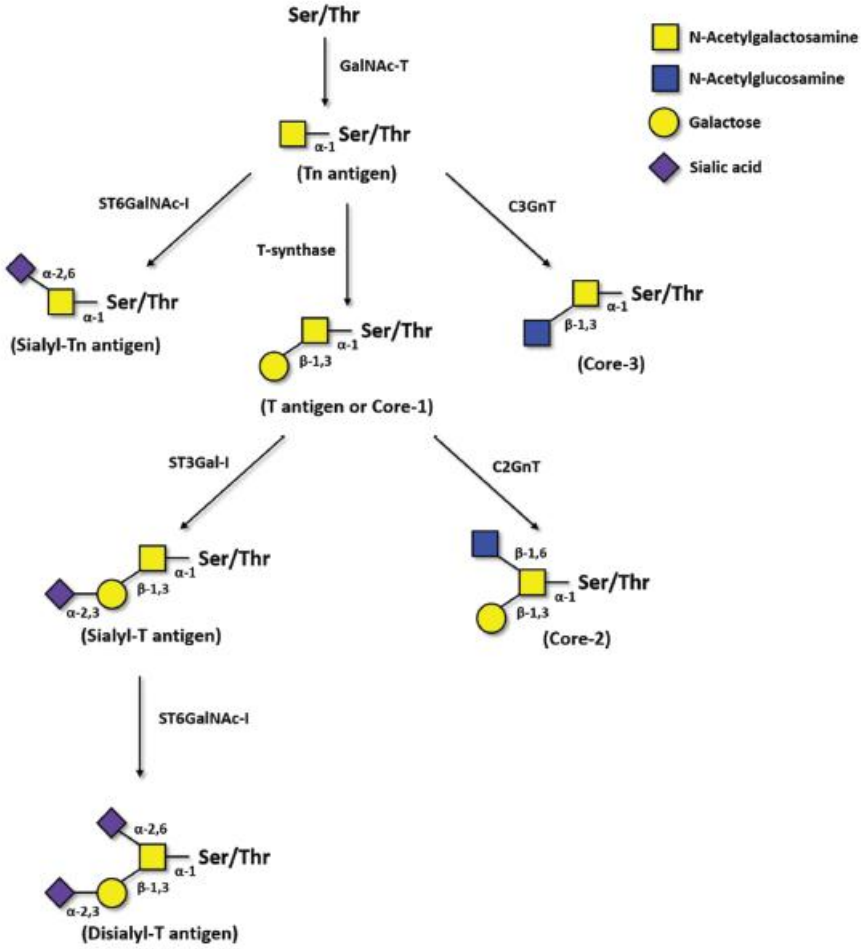


Figure 1.16: Structure of Tn, Sialyl Tn/T and T antigen. This image depicts the shortened glycans known as Tn, sialyl Tn and T-antigen. These glycans are often linked with various types of cancer. Image taken from review by Loureiro et al. (2015).

Breast cancer is one of the most common cancers worldwide and has the highest instance of new cases in women each year. There have been many studies carried out to identify specific causes for breast cancer and a lot of different ways to identify it as early as possible. There are three main groupings of breast cancer, these are; oestrogen or progesterone receptor-positive (ER+), HER2-positive and Triple negative (negative for oestrogen, progresterone and HER2). Tn antigen has been associated with certain types

of breast cancer and is used as a biomarker to try and highlight metastasis and invasive potential (Brooks and Carter 2001). There has been a difficulty in creating a quick, accurate and cost effective technique to identify the presence of any of the Tn and T epitopes. In section 1.3, CD44 was mentioned as a specific cell surface glycoprotein that is involved in a number of cellular processes. Its importance in cell-cell and cell-matrix adhesion makes it an ideal candidate to study when looking at tumour metastasis. CD44 is found to be over expressed in a number of cancers including breast cancer. It's overexpression in cancer can be linked to its ability to cause disruptions in adhesion or promote cell adhesion in new locations during metastasis. In breast cancer, along with a number of other cancers, CD44s glycosylation has been shown to be altered and possess the sialyl-Tn (sTn) and sT antigens (Hauselmann and Borsig 2014). The presence of these antigens, as well as the increased number of CD44 on the surface, is associated with high metastatic potential and a poorer prognosis.

Another important area that has been investigated in breast cancer metastasis is the *N*-acetyl-glucoaminyl transferase 2 (GCNT2) pathway. In metastatic breast cancer tumour cells there has been shown to be an upregulation of the GCNT2 gene. The changes this causes for the glycans on the surface of the cells provokes cell detachment from the extra cellular matrix and promotes cell migration and invasion (Kolbl et al. 2015). Different types of breast cancer can exhibit different glycosylation patterns on the surface. For instance there is different glycans found on the cell surface of ER+ and ER- breast cancers. In a 2011 study it was shown that sialyl-Lewis antigen x (sLe^x) was overexpressed in all ER- breast cancers (Julien et al. 2011). Sialyl-Lewis antigen x is a tetra-saccharide consisting of GlcNAc, fucose, galactose and sialic acid. The structure for sLe^x can be seen in figure 1.16. This increase in surface sLe^x antigen can be accounted for by measuring specific glycotransferase genes that are upregulated in ER- breast cancers. Three enzymes found to be overexpressed in ER- cells are ST3 beta-galactoside alpha-2,3-sialyltransferase 6 (ST3GAL6) and fucosyltransferase 3 and 4 (FUT3, FUT4). These genes being overexpressed facilitates the increased amount of sLe^x found on the cells. Both increased and decreased sialyation of terminal glycans can

have a significant effect on the proteins ability to carry out its function. The sLe^x glycan is found on healthy leukocytes but is now shown to be expressed in cell types not synonymous with this glycan pattern. ER- breast cancer cells can be partially identified using sLe^x as a biomarker but some ER+ cells have been shown to also express this antigen. These ER+ tumours showed increased bone metastasis when compared to ER- tumours expressing sLe^x antigen (Julien et al. 2011). Less than a quarter of high sLe^x antigen expressing ER- tumours showed bone metastasis while essentially all ER+ tumours with high sLe^x antigen expression showed bone metastasis. For these two reasons the presence of this glycan can be used to help identify ER- tumours while also alluding to the metastatic potential of ER+ tumours.

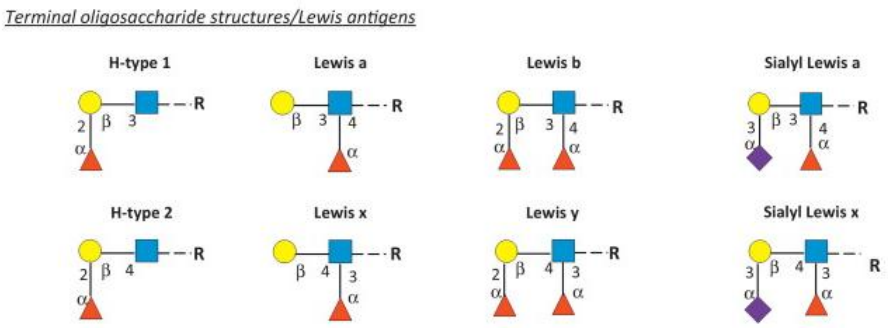


Figure 1.17: Structures of Lewis antigens. This image shows the tri- and tetra-saccharide glycans in the lewis antigen family. Image taken from Pinho et al. (2013).

1.5.3 Glycosylation alterations in colorectal cancers

Section 1.5.2 highlights the changes observed in various types of cancer. In males colorectal cancer (CRC) represents the third most commonly diagnosed cancer, while it also represents the second most commonly diagnosed cancer in females. Statistics supplied by the WHO from 2012 show that there are over 1.3 million cases of newly diagnosed colorectal cancer in the year prior with nearly 700,000 deaths related to the disease (World Health Organisation 2017). CRC is one of the leading causes for cancer related mortality. Even though the majority of deaths are registered from third world countries, where diagnosis is often carried out when the disease has progressed to stage III or stage IV, CRC is still one of the leading causes of cancer-related deaths in the developed world. For instance, in Ireland there are approximately 2,500 newly diagnosed cases each year and it accounts for the second most common cause of cancer related death (Irish Cancer Society 2017). There are various causes of colorectal cancer, some being hereditary while some are caused by lifestyle factors, i.e. dietary or alcohol related (de Freitas Junior and Morgado-Diaz 2016). CRC progression occurs through a series of histopathological stages. These stages range from single lesions, through small polyps often to malignant cancers. As with all cancers early diagnosis is pivotal to ensuring effective treatment. The five year survival rate for CRCs diagnosed in stage I or II are very high, between 80-90%. As the stages progress the five year survival rate drops significantly. Once the disease has progressed to stage III this usually means the CRC has become invasive and often metastases in other organs. In figure 1.18 the five year survival rate for stage III is approximately 63% while the stage IV survival is as low as 7% in males. These statistics show a need for earlier diagnosis and preventative measures to try and stop metastasis of the cancer.

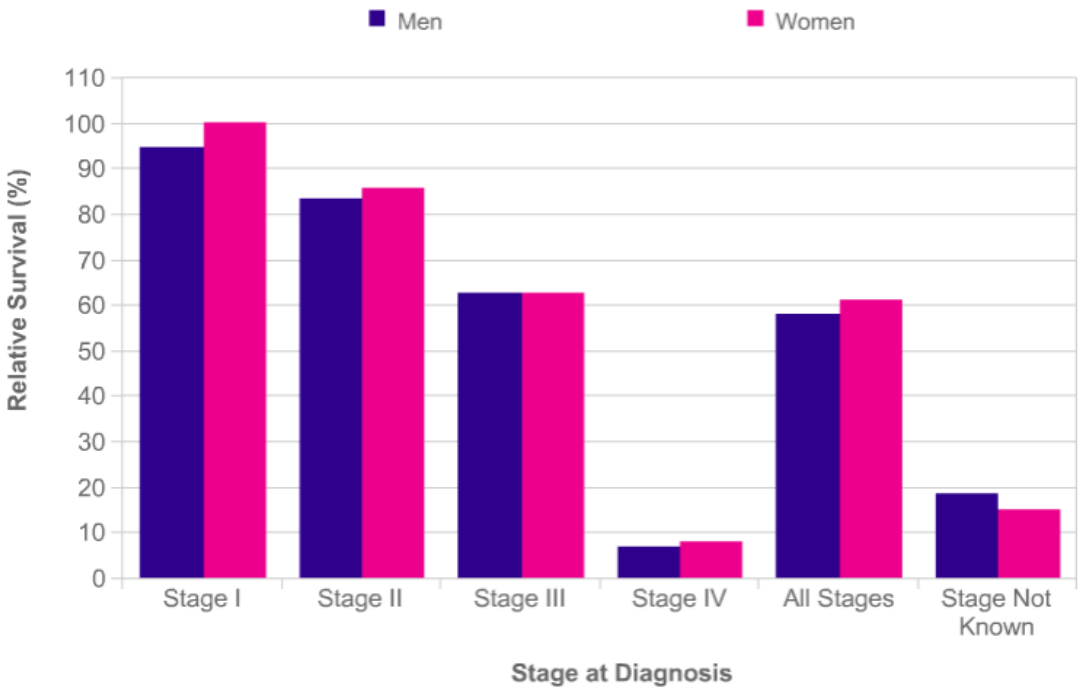


Figure 1.18: Five year survival rates for men and women at different stages of colorectal cancer. This image shows the decreasing survival rate as the CRC progresses from stage I to stage IV. Image taken from Cancer Research UK website (Cancer Research UK 2016)

As with prostate and breast cancers there, has been a lot of research carried out in regards to the cells ability to carry out correct glycosylation. Glycosylation has been shown to be extremely important in epithelial cells ability to correctly adhere to each other. One of the main glycoproteins on the surface of epithelial cells is E-cadherin and numerous studies have suggested that altered N-glycosylation of this protein contributes to the cells inability to correctly adhere. E-cadherin has four N-glycan sites which appear to be involved in calcium dependent cell-cell adhesion at amino acid positions 554, 556, 618 and 633 (Zhao et al. 2008). Incorrect glycosylation at any of these sites can play a crucial role in the cells ability to adhere and can lead to the cells ability to

become metastatic or invasive. Two specific changes to the N-glycans of this protein can affect the adhesion differently. Increased β-1,6- branched N-glycans destabilises the cell-cell binding and consequently aids in tumour progression. This increase is catalysed by an overexpression of the mannoside acetylglucosaminyltransferase 5 (MGAT5) gene (Pinho et al. 2009). There is another MGAT gene that leads to alterations in the N-glycans present on the E-cadherin protein (MGAT3). MGAT3 is the glycotransferase gene involved in β-1,4- mannosyl linkages which leads to increased branching and bisecting of glycans. The regulation of both of these genes has profound roles in many different types of epithelial cell cancers ability to adhere and in their tumour progression. They show ability to effect the stability of cell-cell adhesion as well as cell-matrix adhesion, i.e. they also affect the cells ability to adhere to the extra cellular matrix. The family of glycoproteins affected in the case of cell-matrix adhesion are integrins. Depending on the expression of these transferases there can be increased invasive potential, increased cell migration as well as the affects to cell adhesion (de Freitas Junior and Morgado-Diaz 2016). The changes in E-cadherin mediated cell-cell adhesion of human colorectal cancer cells has been studied and these changes, while found in other epithelial cell cancers, have proven to be involved in the CRCs ability to destabilise cell-cell adhesion. Adheren junctions (AJs) stability in CRCs are heavily affected by altered N-glycosylation and these alterations have been associated with CRC progression (de Freitas Junior et al. 2011).

In the majority of studies to date these changes have been monitored using a mix of mass spectrometry, microscopy and gene expression analysis. Lectins have the ability to monitor some of these changes in a cheap, quick and effective way. Increased branching, change of alpha- and beta- linkages as well as incorrect mono/oligosaccharide addition/subtraction can be identified by using specific lectins to monitor these changes. In the case of colorectal cancer, there has been a lot of research into specific glycan changes that allows for the use of lectin staining, either through lectin micro-arrays or fluorescent microscopy, to identify binding. Both micro-arrays and fluorescent microscopy are limited in their quantitative analysis and largely qualitative, so the use of

lectin staining coupled with flow cytometry will give a better understanding of the quantitative differences in cell surface glycosylation between healthy and cancerous cells, as well as different stages of cancer.

1.5.4 Colorectal carcinoma cell lines SW480 and SW620

In research the use of immortalised cell lines is very common. The breakthrough in the 1950's that resulted in the first human cancer cell line to be immortalised was of an ovarian cancer sample taken from a middle aged woman named Henrietta Lacks (Masters 2002). The subsequent cell line was called HeLa and is still used routinely in research to this day. This breakthrough opened the field to identify and characterise a large number of different cancer cell lines and essentially immortalise a variety of different types of cancer. In the 1950's there were many misconceptions about what cancer was and it was even hypothesised that a series of viruses were the cause of many cancers. The breakthrough of culturing a reproducible human cancer cell line revolutionised the study of human cancers.

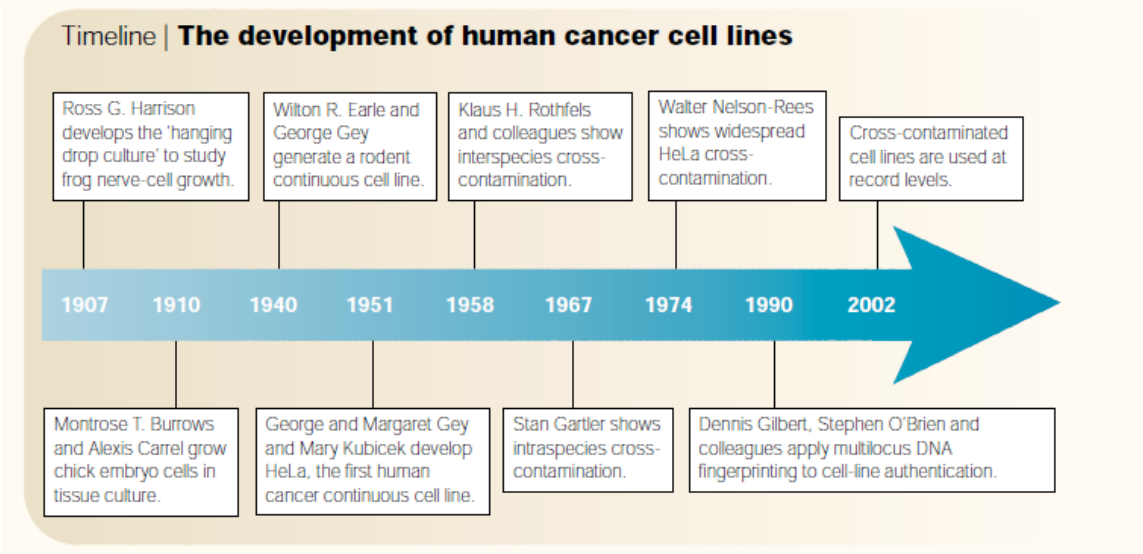


Figure 1.19: Timeline showing the development of human cancer cell lines. This image shows the beginning of modern cell culture and the emergence of reproducible, immortalised cell lines. Image taken from Masters (2002).

Two cell lines that were immortalised were the colorectal carcinoma cell lines SW480 and SW620. These cell lines were originally derived from the same individual one year apart. The fact these cell lines were originally cultured from the same person has made them a unique research tool and the perfect paired cell line. These cell lines were originally cultivated and characterised in the 1970s and have been used in research ever since (Leibovitz et al. 1976).

The SW480 cell line was derived from a primary colorectal cancer. The donor was a 50 year old Caucasian male who had been originally diagnosed with a Dukes' type B colorectal adenocarcinoma, which is defined as an invasion through the bowel wall penetrating the muscle layer but not involving the lymph nodes. The Dukes' classification of colorectal cancers is less commonly used in modern times and the SW480 cell line can be classed as a stage 2 colorectal carcinoma. As a stage 2 carcinoma

it is characterised as a tumour that has not metastasised yet and the cancerous tissue is confined to a single area. This cell line, before it became cancerous, was originally a colorectal epithelial cell. Epithelial cells on the outer lining of the colon/rectum have a certain morphology. These cells must create a barrier to only allow specific molecules pass through the epithelial lining. Due to this the cells will appear to grow in a 'pavement' like fashion, where all the cells adhere to a standard cell culture plate and create a growth where all the cell membranes are directly attached to each other (Peterson and Artis 2014). The SW480 cells have a similar morphology to that of healthy epithelial cells. This morphology can be observed in figure 4.1 in chapter 4.

The SW620 cell line was also derived from a primary colorectal carcinoma from the same patient as the SW480 cell line as mentioned previously. This tissue was harvested from a metastatic site in the lymph node. In the same original study carried out by Liebovitz (1976) this cell line was authenticated and established as the metastatic form of the SW480 colorectal carcinoma. This cell line was originally classed as a Dukes' type C colorectal adenocarcinoma. A Dukes' type C colorectal cancer is believed to have spread through the mucosa layer of the colon and has spread to at least one nearby lymph node. Again, the Dukes' classification system is somewhat outdated and this cancer is now deemed a stage 3 colorectal carcinoma. As a stage 3 cancer it is believed to have metastasised to at least one of the nearby organs, in this cause the lymphatic system. As explained in section 1.5.3, the five year survival rates between the progressive stages of cancer show a decrease in short and long term survival. Morphologically SW620 cells do not exhibit the same phenotype as SW480 cells or healthy epithelial cells. Through the progression of the cancer the cells have taken on a different shape and size and even growth rate. The cells first appear noticeably smaller and more spherical than SW480 cells. The cells appear to become more elongated and can show a more spherical or stretched shape when they have correctly adhered to the plate and reach an adequate confluency. This phenotypic change highlights the change in the cell which led it to metastasise and invade other tissues. This phenotype can be observed in figure 1.19.

1.6 Non-fixed versus fixed cells for cell surface glycan interrogation

Cells that are interrogated using fluorescent microscopy or flow cytometry (Sections 1.6 and 1.7) can be prepared in two ways, fixed and non-fixed. To fix cells, the cells are subjected to a fixative agent which will preserve the cell in its specific cellular state. Fixation causes all cellular biochemical reactions to stop. Common fixative agents used are formaldehyde / paraformaldehyde or glutaraldehyde. Formaldehyde acts as a chemical crosslinker for cellular proteins. The proteins are crosslinked via their free amine groups which preserves the cellular morphology (Hoff 2015). Cells are fixed for a large number of reasons; to stop proteolytic activity of the cells, to allow for analysis over a number of days and weeks using the same sample and to increase cellular permeabilisation. The crosslinking of cell surface proteins can interfere with antigen sites for antibodies due to unnatural surface protein linkages. Cells that are fixed using a chemical fixing agent, such as formaldehyde, are permeabilised and allow for internalisation of certain probes like antibodies.

For analysis of any cell surface antigens or glycans probing non-fixed live cells is recommended (Holmes et al. 2001). To determine the differences between cell surface glycans between cell types using either microscopy or flow cytometry, the cells must not be permeabilised. Cytoplasmic proteins and membranes contain a wide variety of glycans which can also be detected if probing the fixed cells with lectins. Fixation processes that cause protein cross-linking can also alter or change the potential glycan sites on the surface of the cell. Fixation and increased permeabilisation has proven to be a very useful method for intracellular staining with antibodies or other molecules that cannot permeate the cell membrane.

1.7 Fluorescent Microscopy for Glyco-analysis

Fluorescence microscopy is a method that uses fluorescence to study the properties of organic and inorganic substances. With the use of different fluorescent filters, conjugated fluorophores can be excited with the correct wavelength of light and then visualised on the microscope (Galdeen and North 2011). The use of fluorescence in microscopy allows for a user to visualise specific things happening in, and around, a cell. The most commonly used dye for staining a cell nucleus is 4',6-diamidino-2-phenylindole or DAPI. This is a fluorescent stain that binds strongly to regions of DNA that contain high levels of A-T (Tarnowski et al. 1991). It's capable of permeating the cell membrane in live cells and for this reason is used extensively in fluorescent microscopy. DAPI fluoresces in the blue light spectrum. DAPI is routinely used as nuclei counter stain when probing cells with other fluorescently labelled molecules. Cells are then viewed under the different fluorescent channels and the other fluorescent molecules can be detected and visualised (Johnson and Criss 2013). Merged images then can show the fluorescence of different components on one image with a nuclear stain as the control. Cells are either probed with an antibody or other protein with an attached dye or they produce proteins that have been fused to fluorescent proteins (Spector and Goldman 2006).

In respect to glyco-profiling, lectins are often used as the probe and are chemically labelled with fluorophores. As mentioned in sections 1.3 and 1.4, lectins serve as good analytical tools for probing the surface of live cells. In this study where adherent cells are the only cell type being used setting up these lectin probing experiments is relatively simple. Using a DAPI counter stain to visualise the nucleus of the cell, lectins can be used to interrogate the cell membrane of live cells. In figure 1.20 this process of nuclear staining merged with lectin surface staining can be observed (Peiris 2012).

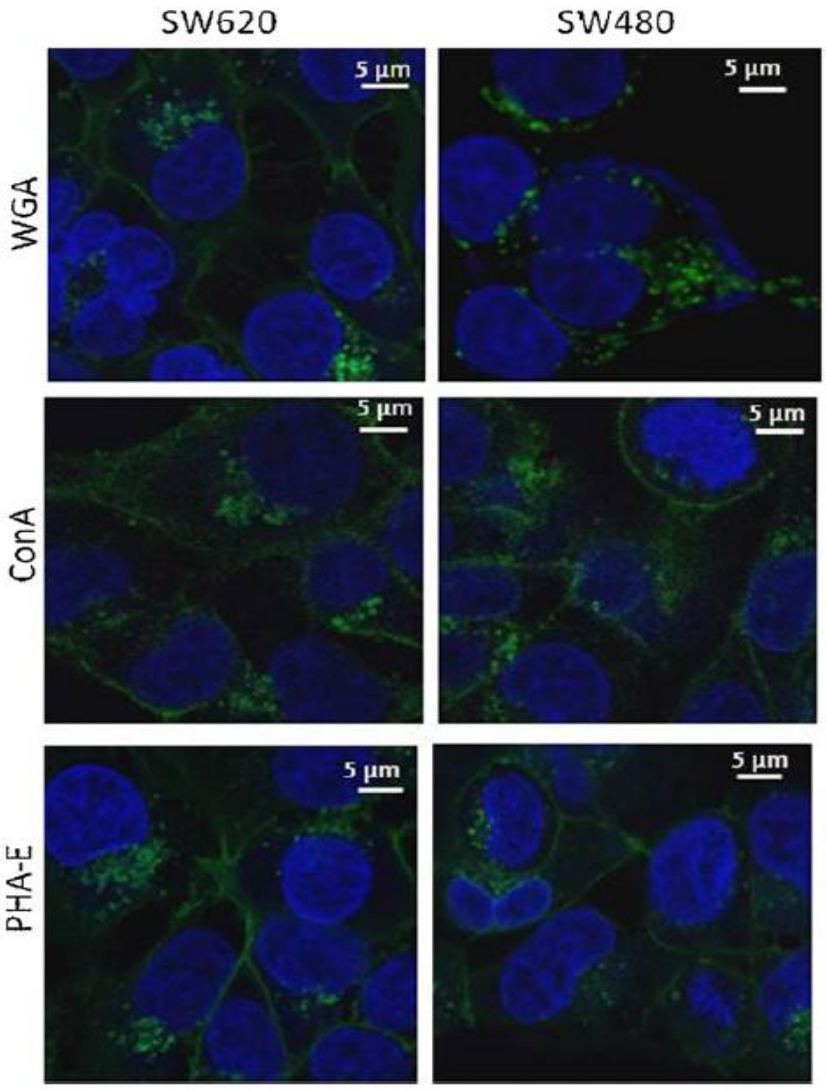


Figure 1.20: Lectin binding to SW480 and SW620 as viewed under a fluorescent microscope. These cells highlight surface lectin binding of WGA, ConA and PHA-E while also showing the DAPI nuclear counter stain. Green fluorescence indicates lectin binding. Blue fluorescence indicates the DAPI nuclear stain. Image taken from Peiris (2012).

The advantages to using fluorescent microscopy as an analytical technique for identifying glycans on the surface of cells are: (i) Glycans can be analysed intact on the cell surface and do not need to be cleaved; (ii) Multiple lectins can be used in the same experiment, i.e. a 12 well plate with cells in each well can have 12 different lectins used on the same cell type; (iii) Competing sugars can be used to selectively inhibit binding of the lectins to show glycan specificity. While this technique has proven to be a very useful tool in glyco-profiling it does have limitations. These limitations are: (i) The method is largely qualitative and the quantification can be highly subjective and (ii) cells grown in the presence of serum can have high background due to the fact a large number of serum proteins are also glycosylated and can attach to the surface of the wells or dishes used for cell culture. Both of these limitations will be further analysed in the discussion of the results section pertaining to the fluorescent microscopy carried out in the project.

1.8 Flow Cytometry for Glyco-analysis

Flow cytometry is a laser-based technology that is used to accurately count cells, sort cells and detect fluorescently labelled biomarkers or probes on cells. This process is carried out on an instrument named a flow cytometer. The flow system is designed to create a constant stream of droplets which each contain a single cell. These single cells are then interrogated by a laser and measurements can be taken based on forward scatter, side scatter and the excitation of fluorescent fluorophores (Ormerod and Novo 2008). Forward scatter is used to measure the size of the cell, whereas side scatter is used to measure the internal complexity of the cell. Both of these parameters can help in determining the health or viability of the interrogated cells and help in distinguishing dead cells or debris from the viable cells. Through the use of different filters and the full light spectrum, a large number of fluorophores can be used to interrogate cells. This process is highlighted in figure 1.21.

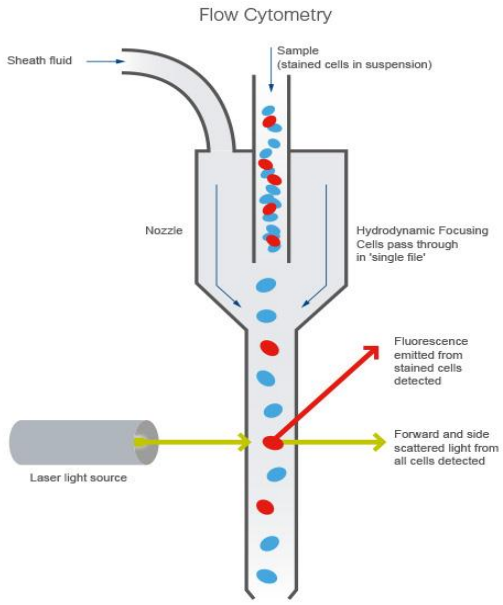


Figure 1.21: Schematic of a Flow Cytometer interrogating cells. A cell suspension is separated into single cells that are interrogated by a laser and specific measurements are recorded. Image taken from abcam.com website (Abcam 2017).

Side scatter and forward scatter can also determine different cells in a cell suspension based on size but if cells are roughly the same size then specific fluorescent probes, such as antibodies or lectins, may be used to identify the different cell groups. The laser excites the specific fluorophores used to probe the cells and these results can be distinguished from the unstained population (Tao et al. 2008). In figure 1.22 the difference between fluorescently positive and negative populations is highlighted. The fluorescent voltage settings are set up against the unstained population as a control. As the technology continues to be modified and enhanced, the number of fluorophores available for purchase and detectable by various cytometers continues to increase. Some fluorophores have excitation and emission spectra that are very close and cannot be easily distinguishable from each other. The ability to compensate for this is one of the more powerful functions of modern flow cytometers. Through the use of specific software this compensation can be applied to the measurements where a large number of different fluorophores are being used on the same type of cells. This compensation allows for the use of multiple fluorophores interrogating cells while accounting for and removing the possibility of false positives. The histograms displayed in figure 1.22 are only using a single fluorophore to detect fluorescence. The use of multiple fluorophores is usually better visualised using scatter or dot plots instead of histograms. An experiment carried out using three different fluorophores to identify three cell populations can be observed in figure 1.23 (Sukhdeo et al. 2013). Antibodies to date have been the most common probes used on cells to be visualised by flow cytometry, but the use of lectins in cell probing has become increasingly routine.

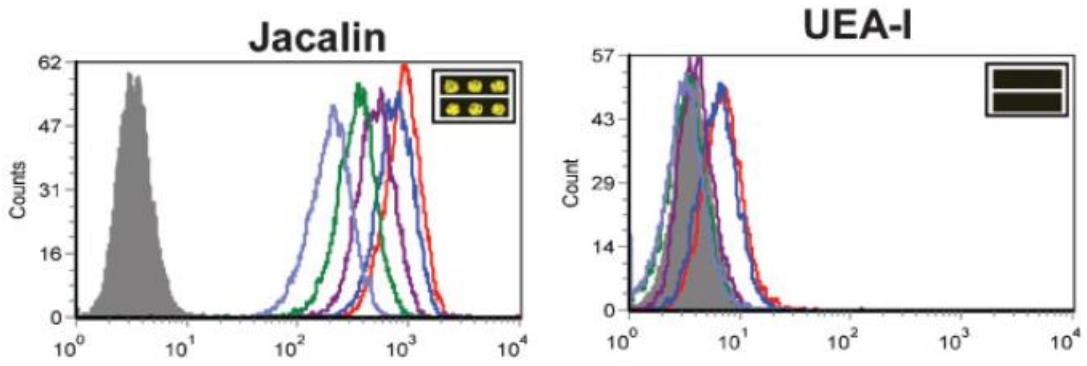


Figure 1.22: Flow cytometric analyses of cells using lectins showing positive and negative binding via histograms. Jacalin at different concentrations binds to these cells while UEA-I at different concentrations does not bind to these cells. The shift in peak relative intensity identifies positive and negative populations. Image taken from Tao et al. (2008).

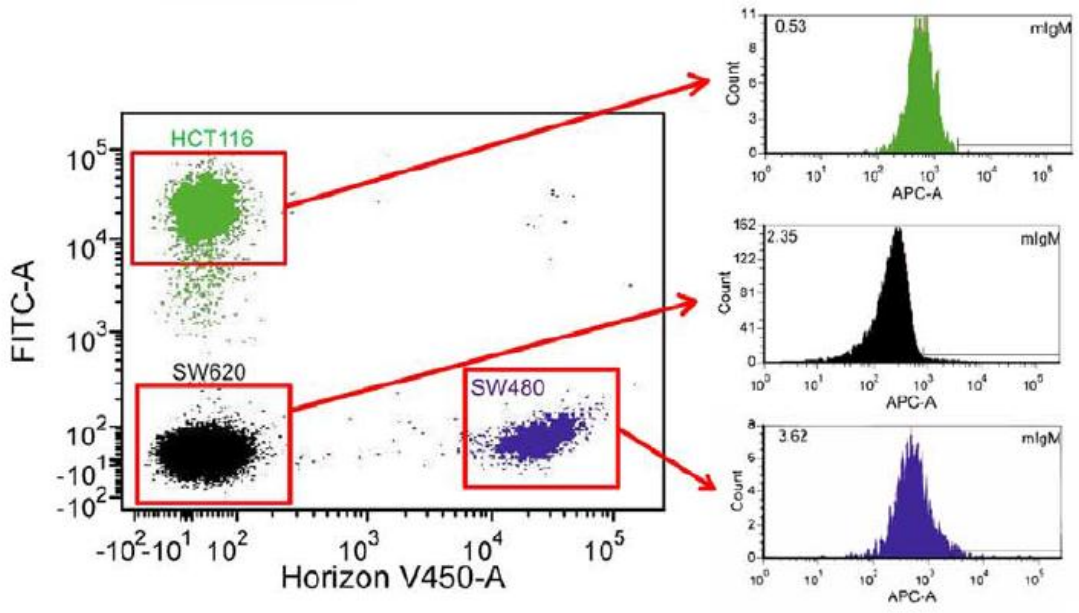


Figure 1.23: Flow cytometric analyses of cells using antibodies to highlight different cell populations in a solution. Three colorectal cancer cell lines are identified by the specific antibody bound to each of them. HCT116 are FITC positive, SW480 are Horizon V450-A positive while SW620 is negative for both fluorophores. All the cells are APC positive. Image take from Sukhdeo et al. (2013).

With regards to glyco-profiling, lectins can be chemically labelled with specific fluorophores and then used to probe cells. This process is similar to the process used in fluorescent microscopy but without the need for a nuclear counter stain. Forward scatter and side scatter are used to determine cell viability rather than the visualisation of an intact nucleus. The advantages of using flow cytometry for lectin probing of live cells are: (i) the cells are live and can be gated accurately to ensure only viable cells are analysed; (ii) a large number of cells can be interrogated in a very short period of time (thousands of cells per minute) to give a large sample size for analysis; (iii) The results are qualitative and quantitative, this allows for relative fluorescence to be calculated and can infer the amount of specific glycans; (iv) Some lectins can cause cytotoxicity and cells that are dead or undergoing apoptosis can be removed through specific gating strategies (Batisse et al. 2004). There are some disadvantages to using a flow cytometer for lectin probing, and these are: (i) Cytotoxicity caused by lectin binding may decrease the number of viable cells in a sample and (ii) lectins may have different levels of fluorophores per individual protein. These two disadvantages can be largely overcome by using intelligently designed gating strategies and limiting variabilities between samples. For these reasons flow cytometry appears to be an underutilised method for glyco-analysis and this project will attempt to prove this can be a useful method for identifying surface glycans in the future.

1.9 Summary

Throughout this introductory chapter the significance of glycosylation has been identified. Cells require correct protein glycosylation to fulfil their desired function. Cells have a large number of glycoproteins embedded in their cell membrane with glycan complexes on the outer surface of the cell. These glycans help to facilitate many different functions cells in different tissues and organs have. Problems then arise when cells cannot correctly glycosylate their proteins and the functions of these proteins get effected. Alterations in glycosylation can occur as a result of mutations in genes associated with glycosylation or through diseases that force cells to change their function, such as cancer. Glycans can be identified and characterised using a number of different methods, some of which can use live cells *in vitro* or others that cleave off glycans for analysis. Alterations in glycosylation have been linked to a large number of different cancers and these alterations can highlight early stage cancer and even infer metastatic potential of certain cancers. In order to accurately diagnose cancerous cells early in the diseased state, the techniques employed to identify and characterise glycans are continually being upgraded and advanced. In the present study, the advancement of lectin based probing platforms has yielded more efficient and cost effective ways to accurately test for surface glycosylation of cells. Lectins have the unique ability to bind specific monosaccharides in complex glycan structures with a large degree of specificity and affinity. With the increasing number of lectins being identified from microbial sources, the ability to regularly produce reliable, accurate prokaryotic lectin probes has continued to increase. Coupled with the eukaryotic lectins that are already commercially available for purchase, comprehensive glyco-profiling can be achieved using lectin based methods such as microarrays, lectin blots, fluorescent microscopy and flow cytometry. Creating a qualitative and quantitative profile of the surface glycosylation of certain cancer cells may lead to rapid identification of cancers at certain stages of growth, as well as identifying invasive potential of the cancer. Throughout the introduction the importance of different areas in glycobiology were explored, these areas included; glycosylation, surface glycosylation, interrogation techniques for determining

glycosylation, lectins both eukaryotic and prokaryotic, cancer and how alterations in glycosylation may be a biomarker for cancer or disease. Through understanding all of these topics, a plan to try and interrogate surface glycans, using recombinantly produced lectins and commercially available lectins, using fluorescent microscopy and flow cytometry was conceived.

1.10 Aims and objectives of this project

The aim of this project has been touched on throughout this introductory chapter. The study of alterations of surface glycosylation in cancerous cells has become increasingly important in recent years. Strong evidence now indicates that alterations in cell surface glycosylation are amongst the earliest changes that occur when cells begin to go cancerous. For research to continue to grow in this area more comprehensive analytical techniques to investigate cell surface glycosylation need to be developed and utilised. Identifying new lectins from prokaryotic sources that have the ability to act as surface probes would add to the library of commercial eukaryotic lectins and increase the potential for lectin based analysis of cell surface glycosylation. Two of the major aims for this project are: (i) to use recombinant DNA technology to alter a prokaryotic lectin, lecB, to possibly alter its binding specificity, affinity, stability and develop an active lectin probe with a fluorescent fusion partner protein and (ii) to use fluorescent microscopy and flow cytometry to analyse cell surface lectin binding and (iii) create a comprehensive glyco-profile for both SW480 and SW620 so as to compare relative binding.

To achieve both of these aims there were a number of technical objectives that needed to be carried out. These objectives were;

- Utilising site directed mutagenesis to create a library of four lecB clones for characterisation.
- Successfully express and purify these lecB variants and characterise their binding potential.
- Identify viable lecB variants for *in vitro* cell glycan analysis.
- Perform qualitative glycan analysis of SW480 and SW620 cells using fluorescent microscopy with a panel of lectins both commercial eukaryotic lectins and recombinantly produced prokaryotic lectins.

- Perform qualitative and quantitative glycan analysis of SW480 and SW620 cells on a flow cytometer again using a panel of lectins both commercial eukaryotic lectins and recombinantly produced prokaryotic lectins.
- Confirm binding specificity by using a competing free sugar to inhibit binding of each lectin while also using a non-competing free sugar to show that it will not inhibit binding.
- Analyse data and compare lectin binding between both SW480 and SW620 to infer glycan differences on the surface of each cell type.
- Highlight the use of recombinant prokaryotic lectins for *in vitro* glycan analysis.

With all of these objectives in place the experiments described in this thesis were designed to achieve project aims in a reproducible and scientifically accurate manner. Thus the overall aim of this work was to develop a new way of investigating cell surface glycosylation using a combination of unique lectin probes and flow cytometry.

Chapter 2

Materials and Methods

2.1 Vectors, Primers, Constructs and Bacterial Strains:

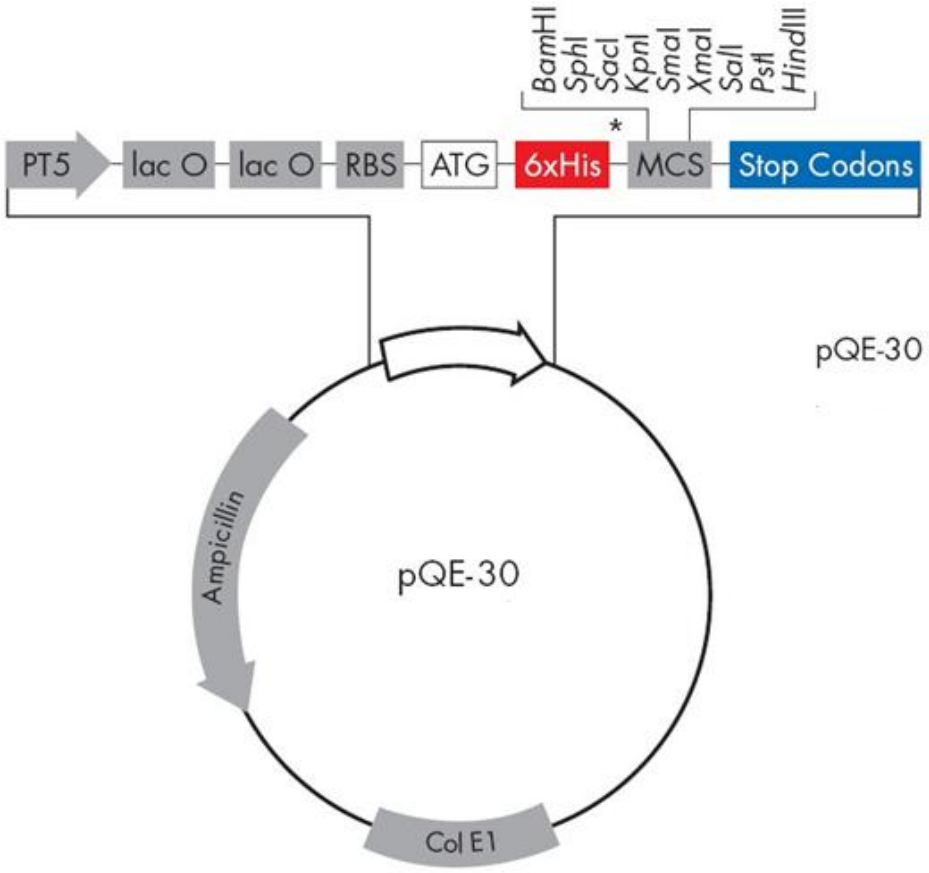


Figure 2.1: pQE30 vector from Qiagen. This 3.4kb DNA plasmid vector was supplied by Qiagen and contains the following features; a multiple cloning site (MCS) preceded by a poly-histidine tag, this His_6 tag allows the protein to be expressed with an *N*-terminal His_6 tag, a beta lactamase gene which allows for ampicillin resistance and a T5 promoter/lac operon for inducible expression of the inserted protein. The vector image was obtained via the Qiagen website (www.qiagen.com).

Table 2.1: Primer sequences

Primer Name	Sequence (5'-3')	Mutation
LecB_S24Tf	GCCAACTCGACCGGAACCCAGACGGTGAAC	S24T
LecB_S24Tr	GAAGGCGGTGACGCCGAACCG	--
LecB_S23A;S24A;G25Af	GCCAACGCGGCCGCAACCCAGACGGTGAACGTG	S23A;S24A;G25A
LecB_S23A;S24A;G25Ar	GAAGGCGGTGACGCCGAACCG	--
LecB_3Kf	AAAAAGAAATCCATGGCAACACAAGGAGTGTTCACC	Three lysine insertion
LecB_Kr	TCCGTGATGGTGATGGTGATG	--

Regions highlighted in blue are the regions containing the mutations.

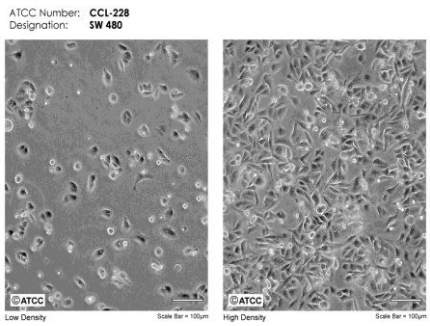
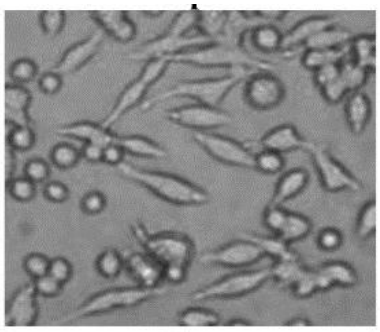
Table 2.2: Plasmid Constructs

Construct Name	Description	Source
pQE30	Expression vector for *N*-terminally tagged His6 proteins, T5 promoter/lac operon, ampR, ColE1 origin	Qiagen
pQE30_eGFP	pQE30 plasmid containing enhanced GFP ORF	ISSC
pQE30_PA-IL	pQE30 plasmid containing PA-IL ORF	ISSC
pQE30_PA-IIL	pQE30 plasmid containing PA-IIL ORF	ISSC
pQE30_PA-IIL_S24T	pQE30 plasmid containing PA-IIL ORF with S24T mutation	This Work
pQE30_PA-IIL_ S23A;S24A;G25A	pQE30 plasmid containing PA-IIL ORF with S23A;S24A;G25A mutations	This Work
pQE30_PA-IIL_3K	pQE30 plasmid containing PA-IIL ORF with an additional three lysine residues between the poly-histidine tag and start codon of the gene	This Work
pQE30_eGFP-PA-IIL	pQE30 plasmid containing eGFP fused to PA-IIL via the *N*-terminus	This Work

Table 2.3: Strains of bacteria used

Strain	Description	Use	Source
E. Coli JM109	FΔtraD36, *pro*Ab+ lacIq, Δ*lac*Z M15 *end*A1 *rec*A1 *hsd*R17 (rk-, mk+) *mcr*A *sup*E44 – *gyr*A96 *rel*A1 Δ (*lacpro*AB) *thi-1*	All-purpose cloning strain that produces stable plasmid DNA at a high copy number.	Sigma
E. Coli KRX	*tra*D36, Δ*amp*P, *pro*A+B+, *lac*Iq, Δ(*lac*Z)M15] Δ*omp*T, *end*A1, *rec*A1, *gyr*A96 (Nalr), *thi-1*, *hsd*R17 (rk-,mk+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), Δ(*rha*BAD) ::T7 RNA polymerase	Protease deficient expression strain. High transformation efficiency.	Promega

Table 2.4: Cell lines used

Cell line	SW480	SW620
Organism	Human	Human
Tissue	Colon	Colon; derived from metastatic site: lymph node
Morphology	Epithelial	Epithelial
Culture Properties	Adherent	Adherent
Biosafety Level	1	1
Disease	Dukes' type B colorectal carcinoma	Dukes' type C colorectal carcinoma
Age / Gender	50 / male	51 / male
Ethnicity	Caucasian	Caucasian
Cell images		
Passage no.	65-70	60-65

2.2 Microbiological and Cell Culture Media

The chemicals and solutions used were all ACS grade and were supplied by Sigma-Aldrich unless otherwise stated. Sterilisation was achieved by autoclaving at 121 °C for 20 minutes. The distilled water (dH_2O) is from a Milli-Q® 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

NaOH is used to adjust the mixture to pH 7.0. This is done before the solution is brought to its final volume. Sterilised by autoclaving (121 °C for 20min.).

*For the production of solid agar plates 15 g/L of Scharlau® bacterial agar was added prior to sterilisation.

Terrific Broth (TB)

Tryptone	12 g/L
Yeast Extract	24 g/L
Glycerol	4 ml/L

The volume is brought up to 900ml and then autoclaved (121 °C for 20min.). Once cooled to room temperature, 100ml of 1M phosphate buffer was added aseptically in order to buffer the solution down to a neutral pH (7.4).

Super Optimal broth (SOB)

Tryptone	20 g/L
NaCl	500 mg/L
Yeast Extract	5 g/L
KCl	2.5mM
pH	7.0

The pH was adjusted using HCl or NaOH depending on the pH of the final solution. This media was then sterilised by autoclaving.

Supplemented RPMI

RPMI 1640, GlutaMAX™	500 mL
Fetal Bovine Serum	50 mL
Penicillin / Streptomycin	5 mL

Penicillin / Streptomycin contains 10,000 units/mL of penicillin and 10,000 µg/mL of streptomycin.

Cell Freezing Stock Solution

RPMI 1640, GlutaMAX™	7.5 mL
DMSO	1.5 mL
Fetal Bovine Serum	1 mL

This solution was stored at -20 °C. RPMI 1640, GlutaMAX™ supplemented and Fetal Bovine Serum were bought from Gibco™. Penicillin and Streptomycin mix were bought from Sigma™.

2.3 Solutions and Buffers:

TE Buffer	
Tris-HCl	10 mM
Na_2-EDTA	1 mM
pH	8.0

1M Potassium Phosphate Buffer	
KH_2PO_4	21.3 g/L
K_2HPO_4	12.4 g/L
pH	7.4

The buffer was sterilised by autoclaving.

RC Buffer 1:	
RbCl	100 mM
$CaCl_2$	10 mM
CH_3CO_2K	30 mM
Glycerol	15% (v/v)
pH	5.8

Made using Ultrapure water. $MnCl_2$ was added after adjusting the pH with KOH accordingly. The solution was filter sterilised through a 0.22 µm filter. The solution was stored at 4 °C.

RC Buffer 2:

MOPS	10 mM
RbCl	10 mM
$CaCl_2$	75 mM
Glycerol	15% (v/v)
pH	6.8

The pH was adjusted using KOH. The solution was filter sterilised through a 0.22 µm filter. The solution was stored at 4 °C.

Phosphate Buffered Saline 10x (PBS)

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

For PBST the detergent Tween-20 was added to give a final concentration of 0.1% (v/v).

Tris Buffered Saline (TBS)

Tris	20 mM
NaCl	150 mM
$CaCl_2$	1 mM
$MgCl_2$	1 mM
pH	7.6

The pH was adjusted to 7.6 with HCl prior to the addition of $MgCl_2$ at a concentration of 1mM. For TBST the detergent Tween-20 was added to give a final concentration of 0.1% (v/v).

SDS-PAGE Buffer (5x)

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

SDS-PAGE Sample Buffer (10x) 8ml

SDS	3.2ml (10% (w/v) stock)
2-Mercaptoethanol	0.8ml
Tris-HCl	2.0ml (0.5M stock, pH 6.8)
Glycerol	1.6ml
Bromophenol Blue	0.5% (w/v)
H_2O	0.4ml

A colorless dye was also made with the same recipe but without the addition of Bromophenol blue.

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 de-stain solution

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

Western Blot Semi-dry Transfer Buffer

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

The volume was brought up to 1 L with dH_2O and stored at 4 °C.

Lysis Buffer

NaH_2PO_4	20mM
NaCl	0.5M
Imidazole	20mM
pH	7.4

*Imidazole concentrations vary from 20mM to 250mM for protein elution using immobilised affinity chromatography (IMAC).

Tris Acetate EDTA Buffer 50x (TAE)

Tris	242 g/L
EDTA (0.5 M)	100 ml/L
Glacial Acetic Acid	57.1 ml/L
pH	8.0

0.7% Agarose Gel

TAE Buffer	1 L
Agarose	7 g/L

Gel was autoclaved to ensure all agarose was dissolved into solution. Various concentrations of 0.7-2% agarose gels were used when necessary. This gel solution can be kept at 60 °C for up to 2 weeks for re-use.

Agarose Gel Loading Dye

Bromophenol Blue	0.25% (w/v)
Xylene Cyanol	0.25% (w/v)
Ficoll Type 100	0.25% (w/v)

Loading dye was sterilised by autoclaving for 20 minutes at 121 °C.

Ethidium Bromide Solution

A 10 mg/mL stock solution of ethidium bromide in dH_2O was stored at 4 °C in the dark in a metal container. For staining of agarose gels, 100 µL of stock solution was mixed into 1 L of dH_2O. The staining solution was kept in a plastic tray, with a plastic top to protect it against light. Once used the ethidium bromide stain was exhausted and the ethidium bromide was extracted by mixing with a de-staining bag (GeneChoice, Gentaur) overnight. The clear liquid was disposed of routinely while the ethidium bromide was incerated. For alternatively staining gels that were used to excise DNA bands, SYBR® safe DNA gel stain (Invitrogen™ - S33102) was used. SYBR® safe DNA gel stain is less harmful than ethidium bromide.

TMB Citrate Solution

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

One 3,3,5,5-Tetramethylbenzidine (TMB) tablet (Sigma T3405) was dissolved in the 10 mL citrate solution to create the TMB solution and 2 µL of H_2O_2 was added directly before use.

2.4 Antibiotics:

Ampicillin was prepared in dH_2O at a concentration of 100 mg/mL and was stored at -20 °C. This was used at a working concentration of 100 µg/mL for solid and liquid media for growth of *E. Coli*.

2.5 Storing of Bacteria:

Bacterial stocks were stored as 27% (v/v) glycerol stocks. 1 mL of a 5 mL overnight culture (late log phase) was aseptically added to 500 µL of sterile 80% (v/v) glycerol. If the bacteria contained a plasmid with antibiotic resistance then appropriate antibiotic was added to the 5 mL overnight culture. Permanent stocks were stored at -80 °C while working stocks were stored at -20 °C. Working stocks were stored on agar plates at 4 °C.

2.6 Culturing of Bacteria:

A loop full of culture from a thawed glycerol stock or a loop from a single colony on a transformation plate was streaked onto a fresh LB agar plate containing the relevant antibiotics. This plate was then incubated at 37 °C overnight for 18-22 hours. Once removed from the 37 °C incubator, the plate can be stored at 4 °C for up to two weeks. Single colonies can be used from this plate to create subsequent plates. A single colony from these plates is used to inoculate 5 mL of LB, with the relevant antibiotic, to create glycerol stocks.

2.7 Isolation of Plasmid DNA:

Plasmid DNA was isolated using a Sigma™ plasmid mini-prep kit (PLN350). The kit was used according to the manufacturer's instructions. The procedure was as follows: 1.5 mL of an overnight bacterial culture was placed in a microfuge tube and centrifuged at 13,000 rpm (rotations per minute) for 2 minutes to pellet the cells. The supernatant is then discarded and the pellet re-suspended in 200 µL of labelled re-suspension solution, this was mixed using a vortex. Once re-suspended 200 µL of lysis solution was added and mixed by inversion. This mixture was left for 5 minutes at room temperature to effectively lyse the cells. After 5 minutes, 350 µL of neutralisation solution was added and mixed by inversion to precipitate all cellular components while leaving plasmid DNA in solution. This mixture was incubated for 5-10 minutes at room temperature. After this time the precipitate was collected by centrifugation at 13,000 rpm for 5 minutes. To prep the spin column, 500 µL of column prep solution was added. This was spun at 13,000 rpm for 1 minute and the flow through was discarded. The supernatant from the cell precipitate was then added to the prepped spin column in a collection tube. This was then centrifuged for 30 seconds to bind the plasmid DNA to the column. Flow-through was then discarded and the column was washed with 750 µL of wash solution and again centrifuged at 13,000 rpm for 1 minute. The flow-through was again removed and the columns were centrifuged again for 2 minutes at 13,000 rpm to remove any excess wash solution and to dry the matrix. The collection tube was then discarded and the spin column was added to a fresh microfuge tube and eluted in 80-100 µL TE buffer (if the samples were being sent for sequencing then elution takes place using dH_2O). This sample was centrifuged for 1 minute at 13,000 rpm and the flow-through contained the desired purified plasmid.

2.8 Agarose Gel Electrophoresis:

DNA was analysed by gel electrophoresis. This was carried out in a BioRad™ horizontal gel apparatus connected to a Labnet Power Station™ 300 power supply. Agarose was prepared by adding an appropriate amount of agar to TAE 1x buffer (see Section 2.3) and was dissolved by boiling at 100 °C (or autoclaving at 121 °C). The levels of agarose varied from 0.7-2% depending on the separation requirements of the DNA. Agarose solutions up to 2% are stored at 60 °C. The agarose solution is poured into plastic trays and allowed to set with a plastic comb fitted to create wells that the DNA samples can be loaded into. The hardened agarose gel was placed in the electrophoresis apparatus with TAE 1x buffer used as running buffer. DNA samples were prepared by mixing them with loading dye using a pipette. Usually 1.5 µL of loading dye is added to 7 µL of sample before loading. Loading dye is coloured so visually it can be seen how far the DNA has migrated through the gel. The glycerol in the loading dye also aids the DNA to sit in the wells and not float into solution. Each well can hold between 10-20 µL of DNA/loading dye sample. Gels were run at 120 V for 20-30 minutes depending on the size of the gel. After completion the gels were stained in an ethidium bromide staining solution (see Section 2.3). Ethidium bromide binds to DNA and can be viewed under UV light. Pictures of these gels were obtained by using a UV transilluminator coupled with an image analyser to capture and print the image. Each gel included a 1kb NEB (New England Biolabs) DNA ladder as a molecular size marker (see figure 2.2). Gels containing SYBR® safe are prepared for DNA extraction from agarose. SYBR safe stain is added to the liquid agarose at a ratio of 1:10,000 and this allows the DNA to be visualised without staining with ethidium.

a) b)

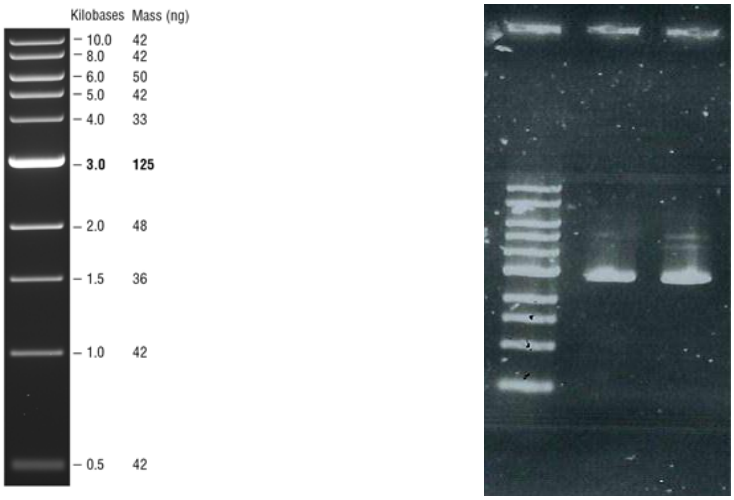


Figure 2.2: DNA size marker. a) The DNA ladder is used as a molecular marker. Image obtained from NEB website (www.neb.com). b) This is an example of plasmid DNA on a 0.7% agarose gel containing a DNA ladder. Lane 1; NEB 1kB ladder, Lane 2; Plasmid sample, Lane 3; Plasmid sample.

2.9 Gel extraction procedure for the isolation of plasmid DNA from agarose gels

To isolate DNA from agarose gels the Gel/PCR DNA extraction kit from GE healthcare® was used. Again this kit was used according to detailed instructions included by the manufacturer. The procedure was as follows: DNA was run on a 1% (w/v) agarose gel pre-stained with SYBR® safe DNA gel stain. The DNA band was excised from the gel using a scalpel. Up to 300 mg of the gel slice was transferred to a microfuge tube. 500 µL of capture buffer was then added to the sample and mixed by vortexing. The sample was then statically incubated at 55 °C for 10-15 minutes (incubation took place in either a thermoblock or waterbath). This sample was mixed

every 2-3 minutes to make sure the gel fully dissolved into solution. Once the sample was fully dissolved it is added to a spin column and collection tube. This spin column was then centrifuged at 13,000 rpm for 30 seconds to bind the DNA to the column. The flow-through was then discarded. 500 µL of wash solution was added and again centrifuged at 13,000 rpm for 30 seconds. The flow-through was discarded and the spin column was centrifuged for 1 minute at 13,000 rpm to dry the matrix and remove any lasting wash solution. This DNA was eluted in 25-50 µL of elution buffer or dH_2O. The elution buffer was added directly to the spin column matrix. This was statically incubated at room temperature for 2 minutes and then centrifuged for 30 seconds at 13,000 rpm. Isolated DNA was stored at -20 °C to insure the DNA integrity.

2.10 DNA sequencing

Recombinant clones and potential mutants were verified by DNA sequencing. Commercial sequencing services were provided by MWG-Biotech/Eurofin genomics. 15 µL of purified DNA (see section 2.7) at concentrations specified by the company was necessary for sequencing. Included with the plasmid DNA were the suitable sequencing primers. Each sample was sent twice, one with the forward primer for a forward reading and the other with the reverse primer for the reverse reading.

2.11 Site specific mutagenesis

Mutations were introduced into the open reading frames (ORF) on plasmid constructs by PCR amplification using complementary phosphorylated primers. The phosphorylated primers carried the designed specific mutation. A standard PCR mixture was set up, as described in section 2.12. The PCR was carried out using Q5 *taq* polymerase. The standard cycle for Q5 *taq* polymerase was used, see Section 2.12. An extension time of 3 minutes was used to amplify 3.5-4 kb plasmids. After introducing a mutation to the plasmid DNA, the template DNA was digested using *Dpn*I restriction endonuclease (See

Section 2.13). *Dpn*I selectively digests the template DNA from any *dam*+ strains of *E. Coli* over the newly synthesised DNA. *Dpn*I is biased towards a methylated recognition sequence. The use of phosphorylated primers protects the new strands of DNA from DpnI. A ligation (see Section 2.14) is then carried out to circularise the newly mutated plasmid. This ligation is facilitated by the phosphorylated primers.

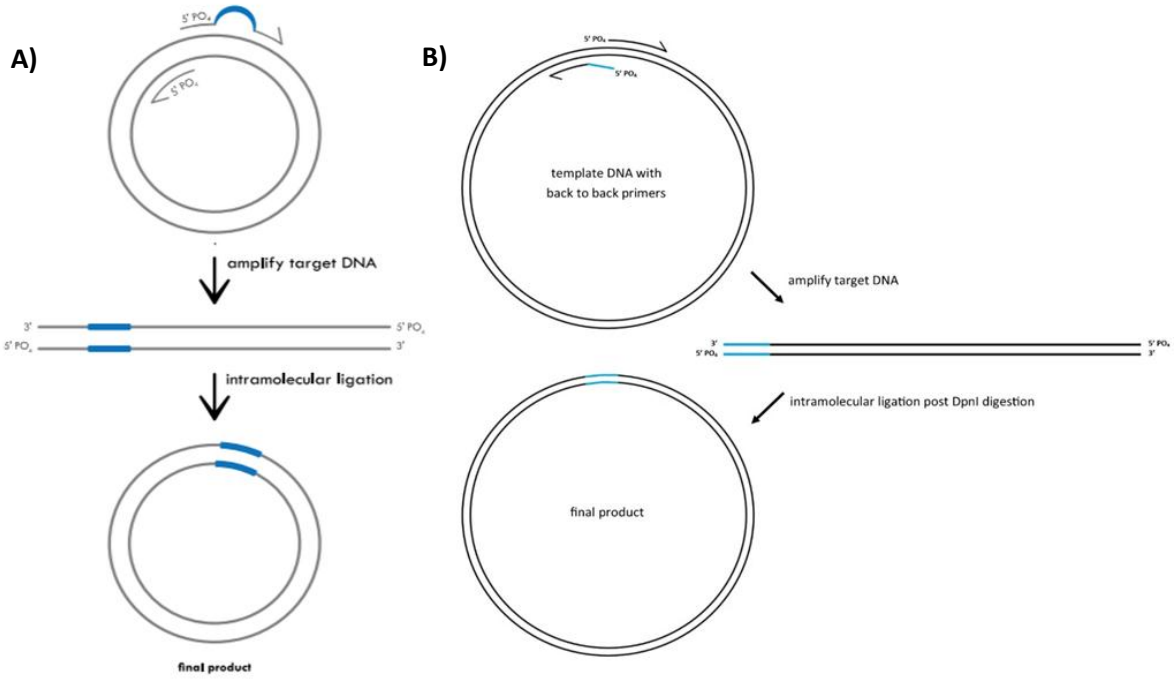


Figure 2.3: Site directed mutagenesis and Insertion through PCR. A) The forward primer contains the desired mutation in blue and does not anneal correctly to the plasmid DNA, B) the DNA to be inserted is added to the 5' region of the primer and through amplification in a PCR mixture, the mutation and insertion take place.

2.12 PCR

2.12.1 Standard PCR reaction mixture:

Template DNA	1 µL	
dNTPs (10µM of each nucleotide)	1 µL	
Primers (10µM)	1 µL	(of each forward/reverse)
Buffer (5X)	10 µL	
dH_2O	35 µL	
DNA polymerase (Q5®)	1 µL	

The final volume of the mixture can vary depending on how much PCR product is required. During this project all PCR reactions were carried out with a final volume of 50 µL. Once the PCR reaction was complete the DNA products wereviewed on an agarose gel. Q5® taq polymerase is a high fidelity polymerase purchased from New England Biolabs®. This ensures that even at low concentrations it will work efficiently. It was standard to dilute the taq polymerase 1:3 before use in the PCR.

2.12.2 PCR programme cycle used for Q5 polymerase reactions

Stage 1:	step 1;	95^{o}C for 10 minutes
Stage 2:	step 1;	95^{o}C for 30 seconds
	step 2;	varying annealing temperatures for 30 seconds
	step 3;	72^{o}C for 30 seconds per kb to be synthesised.
Stage 3:	step 1;	72^{o}C for 10 minutes

Stage 2 was repeated for 30 cycles. Extension time may vary depending on the size of DNA or plasmid being amplified. The annealing temperature was chosen depending on the melting points of both the forward and reverse primers.

2.13 Enzymatic reactions

All the enzymes used and relevant buffers were obtained from Invitrogen Life technologies®, New England Biolabs® or Sigma Corporation and were used according to the manufacturers specifications.

Restriction Digest:

Buffer 10x	5 µL
Restrction Enzyme	2 µL
Template DNA (PCR)	43 µL

The final volume was 50 µL in this case. The reaction mixture was incubated at 37 °C for 120 minutes. All of the reaction is then loaded onto an agarose gels. The correct band is then excised to be recovered using the isolation from gel protocol (see section 2.9). If there is below 43 µL of template DNA, the volume can be adjusted using dH_2O.

2.14 Ligation reaction

During the PCR the use of phosphorylated primers facilitates the ligation of the plasmid DNA (see section 2.11). Using a simple ligation reaction with T4 DNA ligase the plasmid can be re-circularised. T4 DNA ligase and associated buffer were purchased from New England Biolabs® and were used according to the manufacturers' specifications.

Typical ligation reaction:

Template DNA (PCR + restriction)	40 µL
T4 DNA ligase	2 µL
Ligase Buffer (10X)	5 µL
dH_2O	3 µL

Ligation reactions were incubated at room temperature for 3 hours or overnight at 4 °C. This ligated plasmid can then be used in transformations into competent cells. If the amount of template DNA is less than 40 µL, the volume is adjusted using dH_2O.

2.15 Transformations

2.15.1 Preparation of highly Competent Cells

The Rubidium Chloride (RC) method:

This method was first described by Hanahan (1985) and is in brief; 5 mL of LB, with antibiotics as appropriate, was inoculated with a single colony of the relevant bacteria (JM109 and KRX *E. Coli* strains) and left to grow overnight at 37 °C in a shaking incubator at 200 rpm. 2 mL of the overnight culture was then added to 200 mL of LB in a 1 L flask and placed back in the shaking incubator at 37 °C and 200 rpm. The culture was allowed to grow to an OD_{600} of ~0.5 (early-mid exponential phase). Once OD_{600} is reached the flask is removed from the shaker and placed on ice for 10 minutes to cool. All of the following steps took place at 4 °C. The culture was spun down at 5,000 rpm for 5 minutes to pellet the cells. The supernatant was decanted off and the cell pellet was very gently re-suspended in 75 mL of 4 °C RC buffer 1 (see section 2.3). Once re-suspended the suspension was left on ice for 90 minutes. The cells were again collected by centrifugation at 3,000 rpm for 5 minutes. The supernatant was again decanted off and the pellet was gently re-suspended in 10 mL of 4 °C RC buffer 2 (see section 2.3).

The cell suspension was separated into 400 µL aliquots in sterile 1.5 mL microfuge tubes. These were flash frozen using a metal block which had already been cooled down to -80 °C. The competent cells were then stored at -80 °C until use.

2.15.2 Transformation of Competent Cells

A 400 µL aliquot of competent cells were removed from the -80 °C freezer and thawed on ice until fully defrosted. 200 µL of competent cells were mixed with plasmid DNA or ligation reaction. 2-5 µL of DNA was used depending on the concentration in nanograms of DNA. The mixture was then kept on ice for 30 minutes. After 30 minutes the mixture was subjected to a heat shock at 42 °C in a water bath for 45 seconds and quickly returned to ice for 2 minutes. 800 µL of LB broth or SOB media were added aseptically and statically incubated at 37 °C in a water bath for 1 hour. After 1 hour, 200 µL of the mixture was removed and aseptically spread on an LB agar plate, containing the relevant antibiotics if necessary. The remaining 800 µL was concentrated by centrifugation at 13,000 rpm for 2 minutes. The supernatant was discarded and the pellet was re-suspended in 100 µL fresh LB and spread on an agar plate also. The plates were then incubated upside down overnight at 37 °C.

2.16 Standard expression culture

2.16.1 Protein expression for large scale purification

An LB plate with the appropriate antibiotic was streaked from a glycerol stock or single colony from a spread plate of the strain containing the expression plasmid. These plates were streaked for single colonies and incubated upside down overnight at 37 °C. One colony was then selected to inoculate 5 mL of LB broth containing the appropriate antibiotic. The culture was then grown at 37 °C overnight in a shaker at 200 rpm. While this culture is growing overnight a 1 L flask with 200 mL of TB broth is also incubated shaking at 37 °C overnight to insure the TB is free of contaminants. After both have grown overnight and the TB is proven to be contaminant free, 2 mL of the overnight culture is added to the 200 mL TB, this is a 1% (v/v) inoculation. This is again placed in the 37 °C shaker and allowed to grow for 2-3 hours to allow the OD_{600} reach between 0.4-0.6, confirmed using a spectrophotometer (model number). Once this OD_{600} has been reached the culture is induced using isopropylthio-β-galactoside (IPTG). IPTG was added to a final concentration of 50 µM to induce expression of the desired protein. This culture was then allowed to grow at 30 °C overnight after IPTG had been added. This culture was then added to a sterile centrifuge tube and spun down at 5,000 rpm for 10 minutes to pellet the cells. The supernatant was removed, autoclaved and then discarded. The pellet can be processed immediately or stored at -20 °C for future use.

2.16.2 Small scale protein expression to identify viable colonies

Before growing up a larger culture, the starter colony must be screened for maximum expression. Colonies were re-streaked from a transformation plate for single colonies. 10 colonies were streaked onto 10 separate plates for single colonies. All of these colonies were then tested to calculate protein expression. An overnight culture using TB, 1 M phosphate buffer, ampicillin (100 µg/ml) and IPTG (50 µg/ml) was made to a final volume of 8 mL. These overnight cultures were inoculated with a single colony and incubated at 37 °C overnight. The following equation was used;

$0.7/OD_{600}$ x 300 = volume of culture in µL to be harvested

This equation was used to ensure all samples had the same number of cells. The samples were all run on a 15% resolving SDS-PAGE gel. Due to each sample containing the same number of cells, the protein bands are directly comparable to each other and a specific clone is selected.

2.17 Preparation of cleared lysate for protein purification

2.17.1 Cell Lysis by Sonification

A cell pellet from 200 mL expression culture (see section 2.16) was re-suspended in 80 mL lysis buffer containing 20 mM imidazole, pH 7.6 (see section 2.3). The cells were disrupted with a 3 mm micro-tip sonicator (Sonics & Materials Inc.) using 2.5 s, 40 kHz pulses for 60 seconds. This step was carried out three times. The lysate was then centrifuged at 5,000 rpm for 20 minutes at 4 °C (using a GSA rotor) to pellet any insoluble material. The cleared lysate was then transferred to a fresh container. The lysate was vacuum filtered through a Whatman type 1 filter (0.11 µm). The filtered lysate is stored at 4 °C for protein purification.

2.17.2 Cell Lysis by Cell disruption

A cell pellet from 200 mL expression culture (see section 2.16) was obtained. The pellet was re-suspended in 80 mL of 20 mM imidazole lysis buffer with 0.01% anti-foam. The cells were fully homogenised in this buffer. This solution was disrupted using a constant systems cell disruptor at 15 kpsi (15 kpsi is suggested for *E. coli* disruption). The sample is collected and run through the cell disruptor for a second time to ensure total lysis. After the second disruption, 50 mL of 20 mM imidazole lysis buffer was run through the machine to ensure the entire sample was retrieved. The lysate was then added to a high speed centrifuge tube. The lysate was centrifuged at 13,000 rpm for 40 minutes to pellet the insoluble material. For this project the protein was soluble and was found predominantly in the cleared lysate. The cleared lysate was then filtered through a Whatman type 1 filter (0.11 µm filter). The filtered lysate is stored at 4 °C for protein purification.

2.18 Purification of recombinant protein

Immobilised Metal Affinity Chromatography (IMAC) was used to purify the recombinant proteins as they contained an N terminus poly histidine tag (His_6). The principal behind IMAC is that the protein will form a covalent bond with the metal ions. The metal ions used were Ni^{2+} and this metal has a high affinity for histidine residues. The type of resin used in this project was HyperCell™, supplied by Pall Life Sciences.

2.18.1 IMAC purification

Charged resin (see section 2.18.2) is stored in 20% (v/v) industrial methylated spirits (IMS). The resin was washed with 10 column volumes (CV) of dH_2O. The column was then equilibrated with 10 CV's of lysis buffer containing 20 mM imidazole (see section 2.3). Once the column was prepared the filtered lysate (see section 2.17) was gently added to the nickel-nitrilotriacetic acid resin (Ni-NTA). During this step all the His_6 tagged protein will bind to the resin and the other cellular proteins will be collected in the flow through. Some proteins with high levels of surface histidine may also bind to the resin. The column was then subjected to various wash steps with lysis buffer containing 10 mM, 20 mM, 50 mM and 80 mM imidazole. These wash steps cause any other proteins with weaker, non-specific binding to be washed off. To elute off the recombinant protein various concentrations of imidazole were used. These concentrations varied from 125 mM – 500 mM depending on the protein. A lysis buffer containing 250 mM imidazole was typically used. Five CV's of elution buffer were used to elute the bound protein. The eluted protein was collected in 1 mL fractions. These fractions were tested using SDS-PAGE to identify which fractions have the highest purity and concentration of protein. The fraction (or fractions) with the highest concentration of protein will be buffer exchanged to concentrate the protein in a smaller volume into a buffer which would not damage the protein.

2.18.2 Stripping and Recharging the IMAC Resin

The resin was stripped and recharged every 5 uses to maximise protein retention. The resin was first washed with 2 CV of dH_2O followed by 2 CV of 50% (v/v) ethanol or IMS. The resin was then stripped of any metal ions by washing with 2 CV of 100 mM EDTA, pH 8.0. Any remaining impurities on the resin were removed with 2 CV of 200 mM NaCl and rinsed with 2 CV of dH_2O. This was followed by a 10 CV wash with 30% (v/v) isopropanol over the 30 minutes. The resin was then again washed with 10 CV of dH_2O and recharged with 1 CV of 100 mM $NiSO_4$. Finally, the column was washed with 10 CV of dH_2O and stored in 20% (v/v) IMS.

2.19 Buffer exchange/ Desalting of purified protein

2.19.1 Buffer exchanging samples from imidazole to PBS using VIVASPIN columns

After each fraction was collected using IMAC the samples that contained the highest level of pure protein were buffer exchanged into PBS. This buffer exchange was carried out using a special type of microfuge tube and filter, a VIVASPIN 500 manufactured by Sartorius Stedim. The sample was added in 500-800 µL volumes and centrifuged at 13,000rpm for 7-10 minutes. This step was repeated until all the fractions containing relevant purified protein have been centrifuged through the filter column. Each time the flow through was discarded and the column was placed back inside the collection tube for further centrifugation. 700 µL of PBS was used to wash through and again this is centrifuged at 13,000rpm for 7-10 minutes. This step was repeated three times. PBS washes were used to ensure all imidazole was removed from the sample. Once the sample was washed through thrice, it is re-suspended in PBS and transferred to a sterile 1.5 mL microfuge tube. Protein samples were stored at 4 °C, -20 °C and -80 °C.

2.19.2 Buffer exchanging samples from imidazole to PBS using dialysis tubing

Alternatively to section 2.19.1, VIVASPIN columns are not always the optimal choice for buffer exchanging and dialysis was used as an alternative protocol. To prepare dialysis tubing, strips of dialysis tubing (Sigma D9777) were cut and boiled for 30-60 minutes and then left in 100 mM EDTA at 4 °C for short term storage. Before use each piece of tubing was washed out with dH_2O. The IMAC fractions containing the highest concentration of protein were pooled and added to the tubing and sealed. The sealed tube was then added to a 5 L beaker of 1x PBS and left for 24-48 hours. If a large sample was used in the dialysis tubing, the 5 L of PBS was removed after 24 hours and fresh PBS was added. After buffer exchange has occurred the tubing is removed from PBS and placed on polyethylene glycol (PEG). This PEG step concentrated the sample and reduced the total volume of the sample. The tubing is left on PEG for 1-2 hours. After this time the tubing is cut open and the sample is transferred to microfuge tubes and stored at -20 °C for short term storage or -80 °C for long term storage.

2.20 Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis (SDS-PAGE)

All protein samples were analysed via SDS-PAGE. Various percentage resolving gels were employed. Typical esolving gels were; 10% (v/v), 15% (v/v) or 20% (v/v) acrylamide. A 4% (v/v) stacking gel was also used for the loading of samples. The recipes were as follows;

	10%	15%	20%	4%
30% Acrylamide (mL)	2.5	3.75	5	0.325
H_2O (mL)	3	1.75875	0.5	1.54
1.5M Tris-HCl pH 8.8 (mL)	1.875	1.875	1.875	-
0.5M Tris-HCl pH 6.8 (mL)	-	-	-	0.625
10% APS (µL)	37.5	37.5	37.5	12.5
10% SDS (µL)	75	75	75	25
Temed (µL)	3.5	3.5	3.5	2.5

Temed is the last solution to be added to the mixture. Once temed is added the gel must be poured. All percentages are (v/v).

2.20.1 Sample preparation

To prepare a sample to be run on a SDS-PAGE gel some SDS-PAGE loading dye must be added to the sample. 18 µL of the sample was mixed with 2 µL of 10X SDS-PAGE sample buffer (see section 2.3). The samples were then boiled at 100 °C for 5 minutes. The β-mercaptoethanol present in the loading dye and the boiling of the mixture work in unison to denature and unfold the protein.

2.20.2 Sample Application

15 µL of each sample were added to each well, while one well is retained for the relative molecular weight protein marker (Broad range protein marker, NEB). 15 µL of the ladder was added to the reserved well. If the gel was to be used for Western Blot analysis pre-stained protein markers were used (PageRuler™ Plus pre-stained Protein ladder, 10-250 kDa, Life technologies and Expedeon Runblue prestained dual colour marker, see figure 2.4).

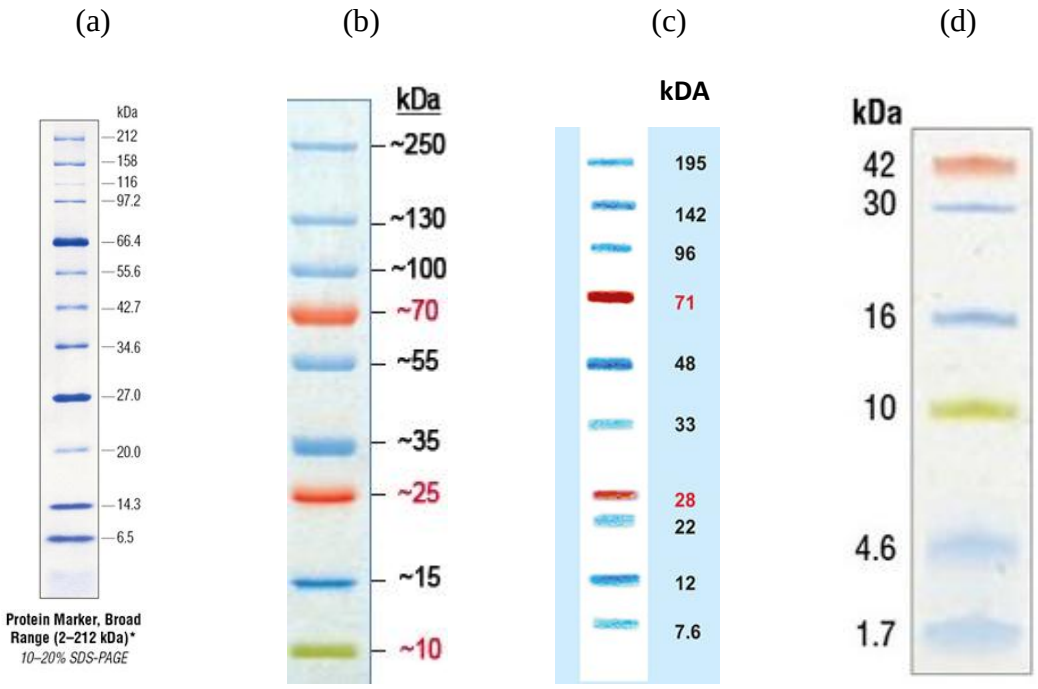


Figure 2.4: broad range protein markers: (a) Broad range protein marker 2-212 kDa, NEB (image obtained from www.neb.com), (b) PageRuler Plus pre-stained protein ladder, Life technologies (image obtained from www.lifetechnologies.com), (c) Expedeon Runblue prestained dual colour marker (NXA05160), (d) Color-coded Prestained Protein Marker, Low Range (1.7-42kDa), Cell Signaling Technology®.

2.20.3 SDS-PAGE gel electrophoresis

Once the wells have been loaded with sample, the gel was run on an Atto AE-6450 vertical gel apparatus connected to a Labnet Power Station™ 300 power supply. Initially each gel was run at 15 mA for 15 minutes to allow the sample migrate out of the well and through the 4% stacking gel. The ampage was then increased to 30-35 mA per gel for 50-75 minutes to allow the sample separate out in the resolving gel. Once the tracking dye has run off the gel the electrophoresis apparatus is stopped and the gel was removed for staining.

2.20.4 Gel visualisation

Polyacrylamide gels were removed from the electrophoresis chamber and first washed with dH_2O. The gels were then stained with Coomassie blue stain solution (see section 2.3) for 1-2 hours. The stain was then removed and the gels were washed again with dH_2O. Coomassie blue de-stain solution (see section 2.3) was then added and left overnight to de-stain the gel. Further de-stain solution may be added if the gel has not fully been de-stained. Once de-stain is removed the gel was washed one more time with dH_2O and then stored in dH_2O and an image was taken.

2.21 Protein Quantification

Protein quantification was carried out using two methods. The first method used was a bicinchoninic acid (BCA) assay. This is a colormetric method where the reagent used will bind to the protein in the sample and a change in colour is observed proportional to the protein concentration. The other method used was quantification by spectrophotometry. Using absorbance readings at 280 nm, the protein can be quantified based on its ability to absorb ultraviolet (UV) light.

2.21.1 Protein quantification using BCA assay

Once the protein samples have been run on an SDS-PAGE gel and buffer exchanged (see sections 2.19 and 2.20), the protein must be quantified. The Pierce™ BCA Protein Assay Kit (Thermo Scientific 23227) was used to accurately quantify the total protein in the sample. The kit was used in accordance with the manufacturer's instructions. The BCA working reagent (WR) was created by mixing reagent A with reagent B (50:1). This assay was carried out using a 96 well microplate (F96 MAXISORP NUNC immune plate) where 200 µL of WR was added to 25 µL of protein sample or BSA standard and repeated in triplicate. The BSA standards were made up in a working range of 20-2,000

μg/mL. The 96-well plate was placed on a shaker for 30 seconds to ensure the contents of the wells were thoroughly mixed and the plate was then incubated at 37 °C for 30 minutes. After 30 minutes the plate was removed from the 37 °C incubator. The absorbance was measured at a wavelength of 570 nm (the suitable range is 540-590 nm) using a BioTek ELx808 plate reader. A standard curve was created using the BSA standards and its linear trend line equation was used to calculate the protein samples concentration.

2.21.2 Protein quantification using 280nm readings

Protein samples were spun down at 13,000 rpm for 10 minutes to remove any debris that may be in the solution. The UV lamp on the spectrophotometer was set to 280 nm and was blanked using the same buffer the protein was stored in. Protein samples were read in triplicate and an average absorbance was determined. The concentration of the protein was then calculated using the Beer-Lambert law;

$$\mathbf{A = \varepsilon cl}$$

Beer-Lambert law used to determine protein concentration using 280 nm absorbance readings. A = absorbance, ε = extinction coefficient of individual protein, c = concentration (in units corresponding to ε) and l = path length (cm).

2.22 Biotinylation of recombinant proteins

The Thermo Scientific EZ-Link® Sulfo-NHS-LC-Biotin kit (21327) was used to biotinylate recombinant proteins. *N*-Hydroxysuccinimide (NHS) activated biotins react with primary amine groups, $-NH_2$, in pH 7-9 buffers and form stable amide bonds. Biotin is a useful label as it is small, 244 Da, and can be conjugated to many proteins without altering their biological activity.

The kit was used according to the manufacturer's instructions. The biotin was removed from the freezer and a 10 mM biotin solution was prepared by adding 180 µL of dH_2O to 1 mg of biotin in a microtube. Equation 3 was used to calculate the amount of biotin to add for a 20-fold molar excess which was then used, in equation 4, to calculate the volume of biotin to add.

$$\text{mL protein} \times \frac{\text{mg protein}}{\text{mL protein}} \times \frac{\text{mmol protein}}{\text{mg protein}} \times \frac{20\,\text{mmol Biotin}}{\text{mmol protein}} = \text{mmol Biotin}$$

$$\text{mmol Biotin} \times \frac{1{,}000{,}000\,\mu\text{L}}{\text{L}} \times \frac{\text{L}}{10\,\text{mmol}} = \mu\text{L Biotin}$$

The appropriate volume of biotin was added to the protein and incubated on ice for 2 hours or at room temperature for 30 min. The sample was buffer exchanged, to remove excess biotin, and concentrated using a 10,000 Da molecular weight cut off (MWCO) spin column, see Section 2.19.

2.23 Enzyme Linked Lectin assay

50 µL of a glycoprotein solution, at a concentration of 5 µg/mL in PBS, was added to the wells of a NUNC Maxisorp 96 well plate. The plate was incubated at 4 °C overnight to allow protein bind to the plate. The unbound glycoprotein was removed by inversion and tapping onto blotting paper towel, then the plate was blocked using 150 µL of a 0.5% polyvinyl alcohol (PVA) solution for 2 hours at 25 °C. The plate was then washed four times using TBST. 50 µL of 5 µg/ml lection solution, diluted in TBS containing 1 mM $CaCl_2$, was then added to the plate and incubated for one hour at 25 °C. The plate was then again washed four times with TBST. 50 µL of a HRP-labelled murine anti-biotin or anti-histidine antibody, diluted to 1:10,000 in TBST, was added to each well. The plate was further incubated for one hour at 25 °C. Following this incubation step the plate was washed four times with TBST. The assay was developed by adding 100 µL TMB solution to each well and left to react for 3 - 10 minutes. The reaction is then stopped by adding 50 µL of 10% H_2SO_4. The absorbance was measured at 450 nm using a BioTek ELx808 plate reader. Each lectin was probed in triplicate and an average was taken over the three wells to get the final absorbance reading. The standard deviation of these triplicates is then calculated and used to create error bars and the ELLA data were presented in a bar chart format.

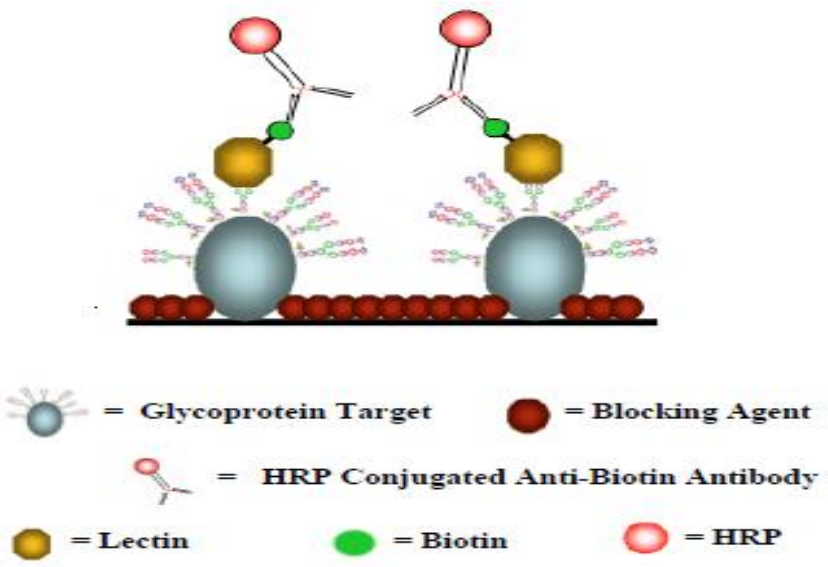


Figure 2.5: Schematic of an ELLA. The ELLA follows the same principal of an anzyme linked immunosorbent assay (ELISA). The glycoprotein target is placed on the surface of the plate and probed with a labelled lectin which is detected colormetrically at a wavelength of 450 nm. The image was obtained from Thompson (2011).

2.24 Lectin specificities

Lectin Name	Specificity
AAL	α-1-6-Fucose
UEA I	α-1-2-Fucose
Con A	Core mannose
RCA	Terminal β-galactose / lactosamine
Jacalin	T-antigen / β-1,3-N-acetylgalactosamine
GSL I	α-N-acetylgalactosamine / α-galactose
GSL II	Terminal GlcNAc
WGA	GlcNAc, NeuNAc
MAL I	Galactose ((β-1,4) N-acetylglucosamine
MAL II	α-2,3-NeuNAc
SNA	α-2,6-NeuNAc
ECL	Galactosyl (β-1,4) N-acetylglucosamine
SBA	Terminal GalNAc
HPA	GalNAc, Tn antigen, GlcNAc
DBA	Terminal GalNAc
GNL	Terminal mannose / high mannose
PNA	β-1,3-galactose
LCA	α-linked mannose
NPL	α-linked mannose, α-1,6-mannose
DSL	β-1,4-GlcNAc (oligomers preferred)

Table 2.5: Specificities of eukaryotic lectins. All commercial lectins were purchased from Vector Labs. Vector Labs supplied information for what buffers and ions were needed for each individual lectin, as well as stating the free sugar need to inhibit binding and the required concentration to inhibit.

Lectin Name	Specificity
LecA (PA-IL)	α-galactose
LecB (PA-IIL)	Fucose, mannose
AAL-2	GlcNAc

Table 2.6: Specificities for prokaryotic lectins. As generated in this project.

2.25 Routine cell culture, seeding/ reseeding cells and cell stocks

2.25.1 Taking Adherent Cells from stock

A vial of the desired cells was removed on ice from liquid nitrogen storage. An aliquot of supplemented RPMI media (Section 2.2) was incubated until its temperature reached 37 °C prior to use. 1 mL of the cell stock is added to 4 mL of pre-incubated media. This cell mixture was centrifuged at 300 *g* for 5 minutes to collect a cell pellet. The supernatant was removed and the cells were re-suspended in 15 mL of supplemented RPMI. This 15 mL cell suspension was added to a T-75 flask (TC Flask T75, Cell+, Vented Cap, Sarstedt) and placed in a 37 °C static incubator to allow cells to grow and proliferate. SW480/SW620 cell lines are between passage numbers 60-70. Cells were passaged a maximum of 15 times before a new stock was used.

2.25.2 Culturing, reseeding and stocking Adherent Cells

Mammalian cells were cultured in supplemented RPMI media (see section 2.2). Media, Trypsin-EDTA (Trypsin-EDTA solution, Sigma) and sterile PBS were pre-incubated at 37 °C for 30 minutes prior to use. A flask containing cells at 75-90% confluency was removed from the incubator, sprayed with 70% IMS and placed in the sterile laminar flow hood. The spent media was taken off and placed in a waste vial and the cells were washed with pre-incubated sterile PBS for 5 minutes. This wash step removes any dead cells and trypsin inhibitors that may be left over from the spent media. 5 mL of 0.5% trypsin-EDTA was added to the flask and the flask was placed back in the static incubator at 37 °C for 5-10 minutes. Once the cells had visably detatched from the bottom of the flask, the flask was then removed from the incubator, again sprayed with

70% IMS, and placed in the laminar flow hood. 5 mL of supplemented media was added to the flask to inhibit the trypsin-EDTA solution. This mixture was removed from the flask and placed in a sterile 25 mL tube. The cells were collected by centrifugation at 1,000 rpm for 5 minutes. The supernatant was removed and the cells were re-suspended in 5 mL of media. A cell count (Section 2.25.3) was performed and cells were re-seeded, at a concentration of 5 x 10^5 cells in 15 mL of supplemented media, into a fresh T-75 flask. These cells were left to proliferate for 2-3 days before re-culturing and seeding.

To create a cell stock, a T-75 flask of 75-90% confluency was cultured. Once the cell pellet was collected and the supernatant removed, the cells were re-suspended in 5 mL of cell freezing stock solution (see section 2.3). 1 mL aliquots were produced containing approximately 10^6 cells. These stocks were stored in liquid nitrogen.

2.25.3 Performing a Cell Count

A cell count is performed using a cell haemocytometer;

(a) (b)

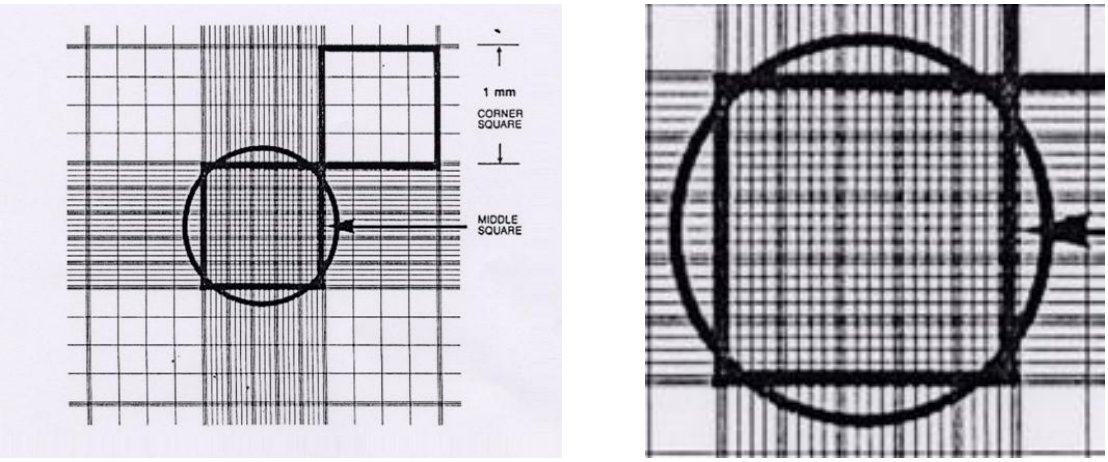


Figure 2.6: Haemocytometer. A haemocytometer is used to accurately count the total cell number in a cell suspension (image obtained from www.phe-culturecollections.org.uk). (a) total haemocytometer, (b) middle section of the haemocytometer.

A cell count is carried out on the re-suspended cells detailed in section 2.26.2. 100 µL of the cell mixture is mixed with 10 µL of Trypan blue (trypan blue solution, Sigma) in a sterile microfuge tube. 20 µL of this mixture is added to the haemocytometer capillary tube and, in unison with the cover slip, the mixture migrates into the chamber. This mixture can now be viewed in the middle section of the haemocytometer under a microscope at 40x magnification. Each smaller grid in the centre chamber is 4 x 4 squares and is part of a larger 5 x 5 square grid. A total cell count of each of the 25 squares was performed in order to calculate accurately the total cell number in solution. Cells with an intact membrane, i.e. live cells, will not allow trypan blue stain their nucleus while any damaged cells encountered will appear blue.

Cell Viability; number of healthy / healthy + unhealthy cells = % cell viability.

Cell Count; middle chamber count x dilution factor x 10^4 = cell number per mL.

Once a cell count has been performed the cells can be re-seeded in a T-75, 6 well plate, 12 well plate, 96 well plate or collected in a microfuge tube for further analysis.

2.26 Fluorescent Microscopy for live cell analysis

Cells were seeded at a concentration of 50,000 cells per well in a 6 or 12 well plate coated for adherent cells (see section 2.25.2). 3 mL of supplemented media containing the cells were added to each well and allowed to proliferate overnight. The following day the 3 mL of spent media was removed and the cells were washed 3 times with pre-incubated sterile PBS to remove all traces of media, which contained serum. The cells were not fixed using any fixative agent. The cells were then probed with lectins (Vector Laboratories®), coupled with/or pre-incubated with a fluorescent dye (dyelight-488), and were incubated for 30-60 minutes. The lectins were used at 5-10 µg/mL. The plates were statically stored at room temperature, in the dark. After the incubation time the cells were again washed 3 times with pre-incubated sterile PBS. Finally 1 mL of PBS was added to each well before the cells were to be viewed under the microscope. Prior to being viewed under the microscope one drop of NucBlue® is added to each well as a nucleic acid counter-stain. NucBlue® Live ReadyProbes® Reagent was purchased from Thermo Fisher Scientific and was used according to the manufacturers' specifications. The microscope used was the Nikon Eclipse Ti. The Nikon Eclipse Ti has three fluorescent filters, FITC (green), DAPI (blue) and TRITC (red). NucBlue® fluoresces in the blue channel, therefore red or green fluorophores are used for the lectins used.

The experiments for fluorescent microscopy were repeated 3 times, with the exception of the inhibiting sugar controls that were repeated twice. The images are purely qualitative and were repeated in order to identify positive and negative lectin binding and not to compare lectin binding between cell types. Each image represents one experiment.

2.27 FluoroFire-Blue ProViaTox Assay

Similar to an MTT assay the FluoroFire-Blue ProViaTox assay measures cell viability. This assay utilizes the redox dye resazurin, which is not fluorescent, but upon reduction by metabolically active cells, a highly fluorescent product is formed. Any living cells can readily reduce this non-toxic reagent and the resulting fluorescent intensity can be monitored using a plate reader.

2.27.1 FluoroFire Blue Assay

A cell count is performed (see section 2.23.3) and the total number of cells per mL was worked out. A serial dilution of cells in supplemented media is plated out, in triplicate, on a 96 well plate coated for adherent cells. The serial dilution begins at approximately 500,000 cells per mL and is diluted 1:2 for each of the first 11 wells. Well 12 contains no cells as a negative control. 200 µL of cells in media at a concentration of 500,000 cells per ml was loaded into well 1. 100 µL of fresh supplemented media was added to each of the corresponding wells. 100 µL from well 1 is removed and added to well 2 and gently mixed. This process was continued until the serial dilution was complete and 100 µL of cell suspension in well 11 was discarded to leave each well containing 100 µL total volumes. Incubate the 96 well plate for at least 1 hour to allow cells to adhere to the plate. Suspension cells do not require an incubation time. After incubation add 20 µL of FluoroFire-Blue solution and incubate at 37 °C for 3-5 hours, in a humidified, 5% CO_2 incubator. An absorbance reading is taken every 90 minutes at two different wavelengths, 570 nm and 600 nm. The resulting data can be plot onto a graph and the cellular viability can be worked out using the equation given by the suppliers.

% Reduction of the FluoroFire Blue ProViaTox assay reagent =

$$\left[\frac{(Eoxi600 \times A570) - (Eoxi570 \times A600)}{(Ered570 \times C600) - (Ered600 \times C570)}\right] \times 100$$

Eoxi570 = molar extinction coefficient (E) of the oxidized FluoroFire Blue ProViaTox assay reagent at 570 nm = 80586

Eoxi600 = E of the oxidized FluoroFire Blue ProViaTox assay reagent at 600 nm = 117216

A570 = absorbance of test wells at 570 nm

A600 = absorbance of test wells at 600 nm

Ered570 = E of reduced FluoroFire Blue ProViaTox assay reagent at 570 nm = 155677

Ered600 = E of reduced FluoroFire Blue ProViaTox assay reagent at 600 nm = 14652

C570 = absorbance of negative control well (media + reagent, no cells) at 570 nm

C600 = absorbance of negative control well (media + reagent, no cells) at 600 nm

2.27.2 FluoroFire Blue assay using Lectin Probes

A serial dilution to determine lectin toxicity on cells was also performed using the FluoroFire Blue assay. The protocol explained in 2.24.1 was followed with this additional step. Starting at a concentration of 100 µg/mL, the lectin was diluted 1:2 through 11 wells, with well 12 having no lectin as a negative control. The lectin range therefore was 0-100 µg/mL. Each well will now contain 5,000-10,000 cells and the serial diluted lectin in supplemented or serum free media at a final volume of 100 µL. 20 µL of FluoroFire Blue reagent was then added to the 100 µl and the equation used in section 2.27.1 was implemented to show cell viability after 3-5 hours.

2.28 Flow Cytometry for live cell analysis

Flow cytometry is the study of cells using a flow cytometer to visualise individual cells. The flow cytometer used is the FACs Aria 1.0. Thousands of cells can be processed through the system per minute, and each cell was interrogated by the lasers individually. The lasers measure forward scatter, side scatter and there are fluorescent lasers that can measure fluorophores at a variety of wavelengths. Forward scatter measures the size of the cell while side scatter measures the internal complexity.

2.28.1 Sample preparation

A cell count is performed (see section 2.25.3) and the total amount of cells per mL was calculated. Cells had been cultured for 48 hours prior to experimentation. Typically, each tube for flow cytometer analysis contained 300,000-500,000 cells. Epindorfs containing the cells were spun down at 300 *g* for 5 minutes to pellet the cells. After the cells have been pelleted, the supernatant is removed and the cells are re-suspended and washed in 200 µL of sterile PBS. The cells are again spun down at 300 *g* for 5 minutes to pellet the cells. The supernatant was removed and the cells were probed with either a fluorescently tagged antibody or lectin probe. All the lectins were biotinylated and pre-incubated with a streptavidin-linked fluorescent dye. The concentration of lectin used was 5 µg/mL diluted in TBS, containing 1mM $CaCl_2$, and the concentration of antibody used is in accordance with the data sheet provided by the supplier. The cells were statically incubated in the dark at room temperature for 20-30 minutes. The cells were again spun down at 300 *g* for 5 minutes and the supernatant was removed. The cells were re-suspended in 500 µL of PBS or TBS and were ready to be processed the flow cytometer.

Four different controls were prepared to accurately assess lectin binding; (i) TBS + 1 mM $CaCl_2$ only, (ii) TBS + 1 mM $CaCl_2$ with DyLight-488 only, (iii) TBS + 1 mM $CaCl_2$ with Dylight-488, lectin and inhibitory sugar, (iv) TBS + 1 mM $CaCl_2$ with Dylight-488, lectin and non-inhibitory sugar.

2.28.2 Sample interrogation and acquisition

The flow cytometer was turned on and set up for acquisition of cells. All the lasers needed for the specific fluorophores used are turned on. The cytometer is set up in accordance with the protocol distributed by manufacturer, FACS Aria 1.0 protocol. Once the cytometer has connected and fluidics is turned on, a new experiment is opened using FACs DIVA software on the associated computer. The parameters are set for the specific samples, i.e. if the FITC channel is only required then remove all other channels from the parameters. The flow rate was to 2 or 3 and the sample was loaded. The voltages for forward scatter, side scatter and the desired fluorescence channel were set. The voltages were set to ensure all cells were visible in the dot plots and histograms that were produced. Dot plots for forward scatter area versus side scatter area, as well as forward scatter area versus forward scatter height were selected. Histograms measuring fluorescent intensity were also selected. Once the voltages were correctly set up and all the cells were be visualized, the sample acquisition was recorded. The amount of cells recorded was 10,000. Experimental files were saved and exported as FCS3.0 files for later analysis.

2.28.3 Flow cytometric data analysis and gating strategy

Flow cytometry data, FCS3.0 files, were opened and analysed using FlowJo. Each FCS file holds the information for a single sample tube. A workspace was created in FlowJo to analyse files from a single experiment or multiple experiments. Within the workspace samples were grouped and sorted based on various attributes, permitting operations to be performed on every sample in a group. The gating strategies implemented were manually applied and followed guidelines set out by Herzenberg et al. (2006).

The gating strategy outlined in figure 2.8 was applied to all data throughout the project. Dead cells and cellular debris can be gated out of the cell population using side scatter area (SSC-A) versus forward scatter area (FSC-A), while the exclusion of cell doublets can be gated out using FSC-H versus FSC-A. Gates were originally constructed around the unstained/unprobed cell populations. This removed dead cells and debris from the sample that has not been probed with anything, i.e. lectins or antibodies. This gate was then applied to all subsequent samples. A histogram was created showing the fluorescence intensity recorded in the FITC-A channel of unstained cells. A single peak was typically observed as these cells have very little to no fluorescence intensity; this histogram was used to separate FITC positive and FITC negative populations. The FITC positive/negative gate was constructed against the unstained cell population, as this unstained population was used as the baseline value for positive and negative intensity. This gating permits the percentage of positive versus negative intensities to be determined. This gate was also applied to all histograms of cells probed with lectins to determine the percentage of positive and negative populations. This can be seen in figure 2.8 images D-F.

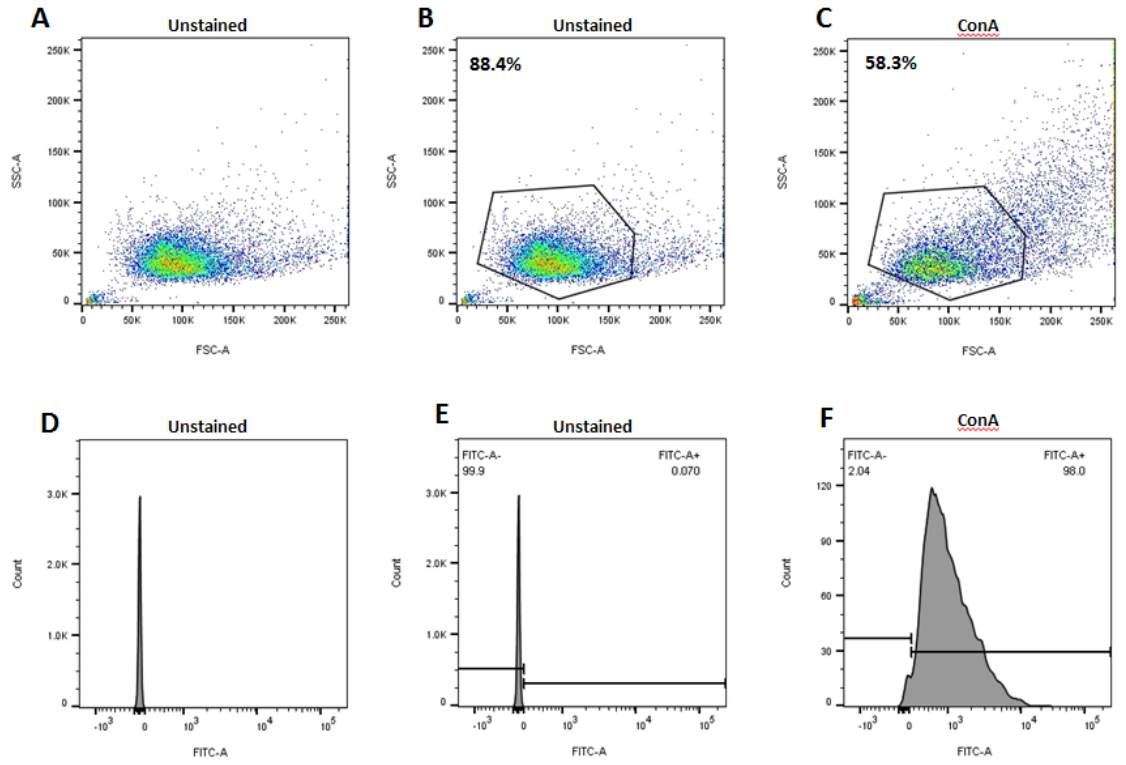


Figure 2.7: Flow cytometry data gating strategy for SW620 cells. A) Raw data, side scatter area versus forward scatter area bivariate plot (dot plot) of unstained cells, B) 'Live' cell gate constructed around cell population of unstained cells, 88.4% of all cells are inside this gate shown on the top left of the plot, C) The live cell gate constructed around unstained cells is applied to cell populations probed with the lectin ConA, 58.3% of all cells are inside this gate shown on the top left of the plot, D) Univariate plot (histogram) of fluorescent intensity (FITC-A channel) of unstained cells in the live gate, E) Histogram of fluorescent intensity (FITC-A channel) of live cells gated for positive and negative populations in the unstained sample, percentage of each population observed in each top corner of the plot, F) The same positive and negative gates from E have been applied to the live gate for cells probed with the lectin ConA, percentage of each population observed in each top corner of the plot.

Once the gating strategy was completed and applied to all samples, a table was generated using FlowJo. This table (table 2.7) shows the percentage of cells within the live gate in a given sample, the percentage of FITC positive and FITC negative cells in a given sample and the overall mean fluorescent intensity value for FITC-A. This data is graphed in a bar chart format to complement histogram plots.

Sample	% in Live Gate	% Negative	% Positive	Mean fluorescent intensity
Unstained	88.4	99.9	0.01	3.7
ConA	46.1	0.98	99	2545
Jacalin	48.4	9.43	90.6	2318
NPL	72.3	0.092	99.9	1185
GSL II	83.6	99.7	0.026	10.5
LecB	49.9	1.22	98.8	800

Table 2.7: Flow cytometry sample data for surface lectin binding to SW480 cells. For each sample, 10,000-15,000 events were recorded in triplicate. This table represents the percentage of cells in the live gate, positive/negative gate and the mean fluorescent intensity of the FITC measure over the whole population in the live gate, i.e. positive and negative cells.

The percentage of events inside the live gate for each sample is indicative of the cytotoxic effect that some lectins have on cells. For unstained cells, they are routinely between 80-90% viability inside the live gate. Some lectins, such as ConA, will significantly reduce the cell number inside the gate (46.1%), where as some lectins, such as NPL, will only have a negliable effect on the cell number inside the gate. This is observed in table 2.7.

2.29 DNA and Protein Databases

A number of online databases and programs were used to analyse DNA and protein sequences.

2.29.1 ExPASy

Expasy allows a nucleotide sequence to be copy and pasted into a window to translate it to a protein sequence. Expasy provides information on each reading frame and indicates which reading frame is correct. Protein sequences can then be run in another program to analyse sequence homology and orthology. The protein sequence can also be used in alignment programs to compare with other sequences.

Weblink: http://web.expasy.org/translate/

2.29.2 BLASTn and BLASTp

Basic Local Alignment Search Tool (BLAST) is a bioinformatics system that can compare DNA and protein sequences with the sequence library database. A BLASTn is used for nucleotide sequences and BLASTp is used for protein sequences. Once the sequence is input it is compared to all sequences in the database and a percentage sequence similarity will be given. The program can search all databases or specific ones, i.e. human only or human, mouse and rabbit. BLAST will analyse sequence homology and orthology with other sequences in the database.

Weblink: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch

2.29.3 Multalin

Multalin is an online alignment program. It can align DNA and protein sequences inputted in a FASTA format. Multalin can align a large number of sequences and also produce a consensus sequence, i.e. two sequences contain the same nucleotide at position 33 and the third sequence contains a different nucleotide at position 33, the consensus will show the more common nucleotide in the reference sequence.

Weblink: http://multalin.toulouse.inra.fr/multalin/

2.29.4 PyMOL

PyMOL is an open source protein molecular visualization system. PyMOL can produce high quality 3D images of proteins from protein sequences. PyMOL can also create 3D models of small molecules and ligands. A protein sequence can be uploaded and a 3D image will form. Once the 3D model is displayed on the system, any section of the sequence can be highlighted and mutated to visualise what changes this would potentially make to the protein.

Weblink: https://www.pymol.org/

2.29.5 Snapgene

Snapgene is a plasmid mapping program that permits the identification of plasmid characteristics. Snapgene allows for *in silico* restriction analysis and permits the generation of primers for specific regions of DNA found on the plasmid. Any plasmid can be inputted into the Snapgene program and the program displays a number of features, i.e. highlight regions of interest with a specific gene or restriction site.

Weblink: http://www.snapgene.com/

2.30 Statistical calculations

2.30.1 ANOVA single factor testing

An ANOVA single factor test was used to identify if there were significant differences between two mean values with a minimum N value of 3. The ANOVA test was carried out in Microsoft excel using the data analysis ANOVA single factor function. The P value limit was set at <0.05 which signified a less than 5% chance of this difference being due to random chance. Any P value >0.05 was deemed to not be significant. All significant differences between two means is denoted with the symbol '*' above the standard deviation error bar. Any P value < 0.001 is denoted with the symbol '**'.

2.30.2 Setting a threshold limit for positive/negative ELLA readings

ELLA (Section 2.23) is a method for detecting lectin binding to a glycoprotein with specific glycans on the surface. To determine if a reading was positive or negative for binding a lower threshold limit of detection was calculated based on the internal PBS control on each individual plate. PBS control absorbance readings varied from 0.05 – 0.1 at OD 450 nm depending on which lectin was used. Due to the variance between these PBS controls for each individual lectin being different, the threshold was set at two times the baseline PBS value is each case. All readings above this threshold limit are denoted with the symbol '†' above the standard deviation error bar.

Chapter 3

Site Directed Mutagenesis of LecB

3.1 Overview

This chapter describes the optimisation, expression and purification of LecB, as well as the site directed mutagenesis carried out on the PA-IIL (LecB) gene. The four mutants that were designed and generated were LecB S24T, LecB with three additional lysine residues (LecB 3K), LecB S23A;S24A;G25A (LecB 3A) and the fusion protein eGFP-LecB. Each variation of the LecB protein fit a specific purpose and this chapter utilises a number of techniques, such as polymerase chain reaction (PCR), restriction digests, immobilised metal affinity chromatography (IMAC) and enzyme linked lectin assay (ELLA), to ensure these proteins were mutated, expressed and purified accurately and that the protein remained active. PCR was used to incorporate the specific mutations. For the eGFP-LecB fusion a number of restriction enzymes are used to cut out the LecB gene and cut the plasmid containing the eGFP to allow for LecB insertion. The LecB gene is then ligated into the open plasmid. DNA sequencing was then used to confirm each plasmid had the correct mutation added and that the fusion gene was generated. Once the plasmids had been confirmed proteins underwent expression checks to identify the clones expressing the most protein and then larger scale cultures allowed for increased protein expression and cell lysis. The cleared cellular lysate can then be used for IMAC purification of proteins. ELLA is a comprehensive assay that allows for the proteins to be validated. Each commercially purchased eukaryotic lectin was tested on ELLA as well as all recombinantly produced lectins. Glycoprotein targets were used to determine binding specificity as well as using competing and non-competing free sugars. The wild type LecB was first compared to other eukaryotic lectins that have similar specificities and then all subsequent LecB mutants produced were tested against wild type LecB for characterisation. Using these methods the proteins may be accurately mutated, purified and characterised before use in *in vitro* analysis (see chapters 4 and 5).

3.2 LecB plasmid purification and DNA sequencing

A previous stock of JM109 *E. coli* containing a pQE30 plasmid with the PA-IIL (LecB) gene was defrosted and streaked for single colonies (see section 2.7). LecB is encoded on a pQE30 vector so the poly-histidine tag was located on the *N*- terminus of the protein rather than the *C*- terminus. The *C*- terminus of LecB is located close to part of the glycan binding region as well as the metal ion binding region of the protein. Any attempt to alter the *C*- terminus effects protein folding and results in failure of the protein to remain active. The pQE30 plasmid contains a T5 promoter and ampicillin resistance gene. These two characteristics allow for the protein to be selectively grown and selectively over expressed. The LecB plasmid was purified and sent for DNA sequencing to ensure the original glycerol stock contained the right LecB sequence. Once the LecB sequence was confirmed and verified using multiple online tools and databases (See section 2.29), the plasmid DNA was ready to undergo site directed mutagenesis. The LecB nucleotide sequence was aligned against the LecB reference sequence obtained using the basic local alignment search tool (BLAST). This alignment was carried out using the online alignment software Multalin and can be viewed in figure 3.1. Multalin gives a consensus sequence to highlight any differences between the compared sequences.

```
           1       10        20        30        40        50        60        70        80
           |--------+---------+---------+---------+---------+---------+---------+---------|
LecBref    ATGGCAACACAAGGAGTGTTCACCCTTCCCGCCAACACCCGGTTCGGCGTCACCGCCTTCGCCAACTCGTCCGGAACCCA
LecB       ATGGCAACACAAGGAGTGTTCACCCTTCCCGCCAACACCCGGTTCGGCGTCACCGCCTTCGCCAACTCGTCCGGAACCCA
Consensus  ATGGCAACACAAGGAGTGTTCACCCTTCCCGCCAACACCCGGTTCGGCGTCACCGCCTTCGCCAACTCGTCCGGAACCCA

           81       90        100       110       120       130       140       150       160
           |--------+---------+---------+---------+---------+---------+---------+---------|
LecBref    GACGGTGAACGTGCTGGTCAACAACGAGACGGCCGCGACCTTCAGCGGGCAAAGCACCAATAACGCCGTCATCGGCACCC
LecB       GACGGTGAACGTGCTGGTCAACAACGAGACGGCCGCGACCTTCAGCGGGCAAAGCACCAATAACGCCGTCATCGGCACCC
Consensus  GACGGTGAACGTGCTGGTCAACAACGAGACGGCCGCGACCTTCAGCGGGCAAAGCACCAATAACGCCGTCATCGGCACCC

           161    170       180       190       200       210       220       230       240
           |--------+---------+---------+---------+---------+---------+---------+---------|
LecBref    AGGTGCTCAACTCCGGCAGCAGTGGCAAGGTACAGGTCCAGGTCAGCGTCAACGGCCGCCCCTCGGATCTGGTCTCGGCA
LecB       AGGTGCTCAACTCCGGCAGCAGTGGCAAGGTACAGGTCCAGGTCAGCGTCAACGGCCGCCCCTCGGATCTGGTCTCGGCA
Consensus  AGGTGCTCAACTCCGGCAGCAGTGGCAAGGTACAGGTCCAGGTCAGCGTCAACGGCCGCCCCTCGGATCTGGTCTCGGCA

           241    250       260       270       280       290       300       310       320
           |--------+---------+---------+---------+---------+---------+---------+---------|
LecBref    CAGGTAATCCTGACCAACGAGCTGAACTTCGCCCTGGTCGGCTCTGAAGACGGCACCGACAACGACTACAACGACGCCGT
LecB       CAGGTAATCCTGACCAACGAGCTGAACTTCGCCCTGGTCGGCTCTGAAGACGGCACCGACAACGACTACAACGACGCCGT
Consensus  CAGGTAATCCTGACCAACGAGCTGAACTTCGCCCTGGTCGGCTCTGAAGACGGCACCGACAACGACTACAACGACGCCGT

           321    330       340     348
           |--------+---------+-------|
LecBref    CGTGGTGATCAACTGGCCGCTCGGCTAG
LecB       CGTGGTGATCAACTGGCCGCTCGGCTAG
Consensus  CGTGGTGATCAACTGGCCGCTCGGCTAG
```

Figure 3.1: LecB nucleotide sequencing results compared to reference LecB sequence. Alignment is carried out using the online alignment tool mutalin. (http://multalin.toulouse.inra.fr/multalin/).

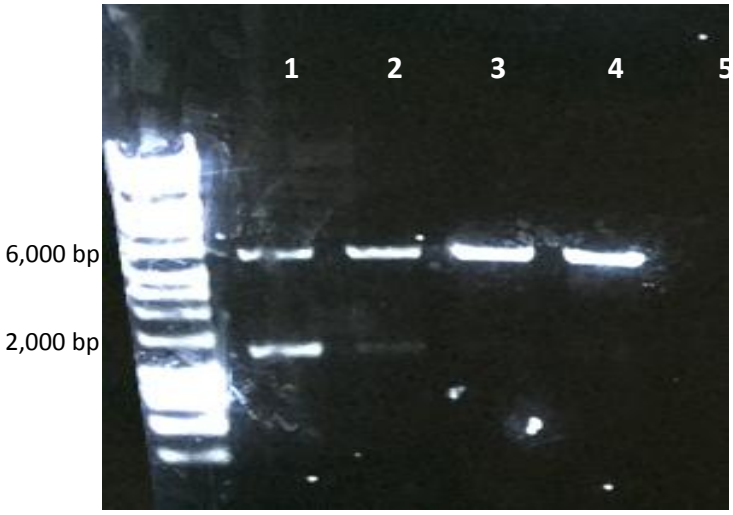


Figure 3.2: Purified DNA plasmids run on a 0.7% agarose gel. Lanes; 1) 1 kb DNA ladder, 2) LecB plasmid, 3) LecB S24T plasmid, 4) LecB 3A plasmid, 5) LecB 3A plasmid.

Purified plasmids from LecB and some of the LecB variants can be seen in figure 3.2. There are 2 orientations plasmids can be found in after mini-prep purification. These orientations are supercoiled and open-circular. Open-circular plasmids run slightly larger than linear DNA of that size whereas supercoiled DNA runs smaller than linear DNA of the same size (Simple Cloning Lab 2017). The pQE30 plasmid with the LecB gene added is approximately 3,800 nucleotides. Both the open-circular plasmid and supercoiled plasmids run at roughly 6,000 bp and 2,800 bp respectively. Through sequence confirmation it is assured that these samples contain the correct LecB variant plasmids.

3.2.1 Designing primers for site directed mutagenesis of LecB plasmid

The wild type LecB plasmid is used as the DNA template for any polymerase chain reaction (PCR) that is used to introduce mutations to the sequence. For site directed mutagenesis the desired mutation must be incorporated into either the forward primer, reverse primer or both in the PCR. There were a number of mutations hypothesised throughout the project that would be either beneficial or detrimental to the final protein product. Preliminary results from past mutagenesis carried out on LecB suggested that a change to the Serine at position 24 would affect the proteins ability to bind its specific sugars. LecB, being a dual specific lectin, binds both fucose and mannose, although it binds mannose to a lesser degree. Originally the design for changing the serine (S) at position 24 to a threonine (T) was to remove mannose binding while keeping relatively strong fucose binding. This hypothesis was driven by the decrease in distance from the hydroxyl group on threonine to the sugar over serine. Both serine and threonine contain hydroxyl groups but in the orientation of this group on threonine at position 24 would change this distance possibly changing the proteins ability to bind mannose. This primer design was hypothesised based on the 3D model of LecB in complex with both mannose and fucose using the 3D modelling software PyMol. Figure 3.3 A shows LecB in complex with fucose. Figure 3.3 B and C show the difference in distance between the hydroxyl groups on the mannose sugar and the native serine amino acid and the threonine residue on the mutated LecB S24T.

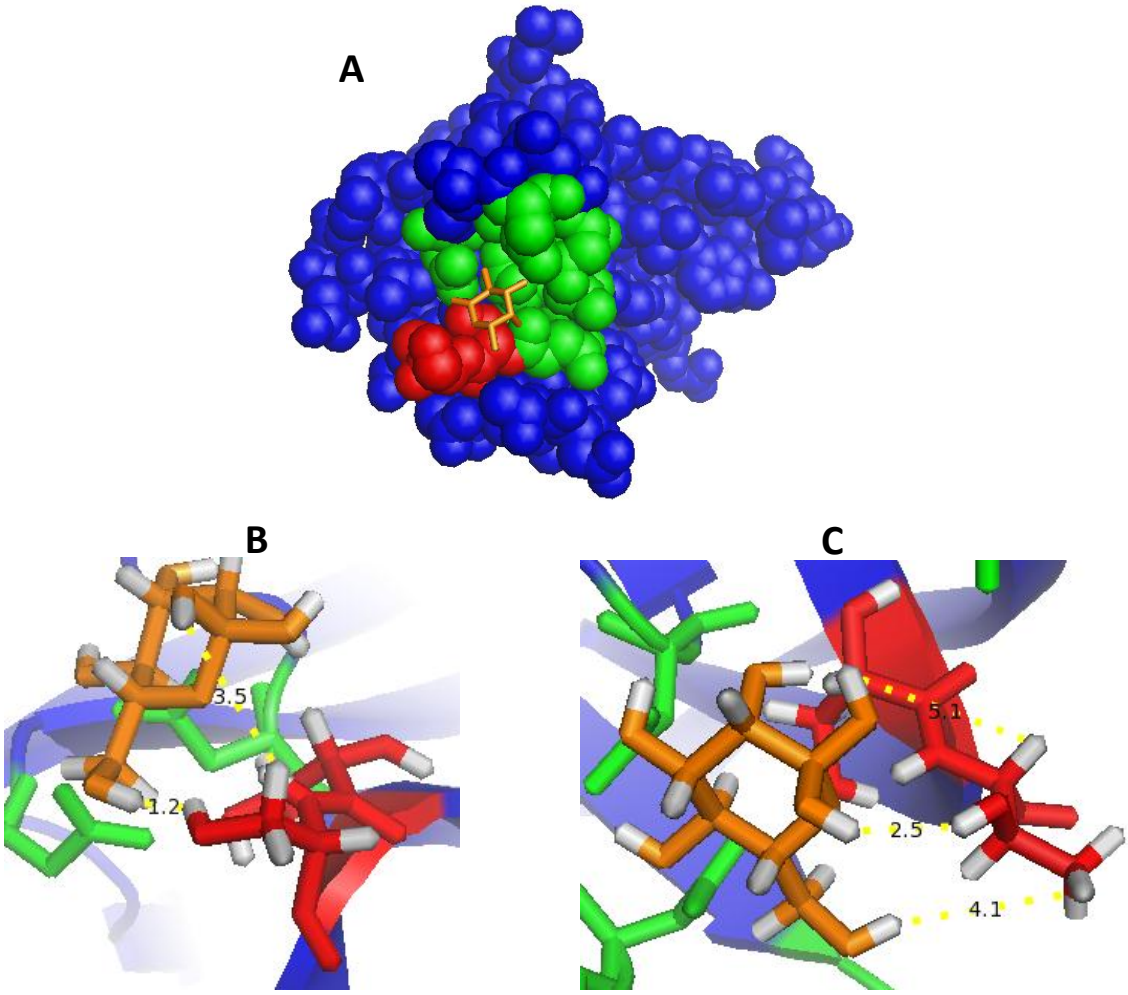


Figure 3.3: LecB protein and sugar interaction structures generated using PyMol. A) LecB protein with fucose sugar molecule in complex, Blue regions denote the protein, green regions involved in sugar binding as well as calcium binding, red regions denote the sugar binding regions highlighted during this project for mutagenesis and orange denotes the sugar molecule. B) Distance between the hydroxyl groups on both the mannose sugar and native serine on the LecB molecule at amino acid position 24. C) Distances between the hydroxyl groups on the mannose sugar and the threonine residue on the mutated LecB S24T. Measurements are denoted using the distance in Ångströms.

The other mutations designed were to add multiple lysine residues between the poly-histidine tag and start codon of the LecB gene, and to replace three of the amino acids involved in binding at positions 23-25 to alanine residues. Alanine scanning is used to replace polar amino acids with non-polar alanine to remove functionality of the protein (Morrison and Weiss 2001). The three residues shown to be involved in binding in LecB are serine, serine and glycine from positions 23-25 respectively. These three residues are involved in glycan binding but not involved in the proteins metal ion binding pocket. The original hypothesis was to allow the calcium binding pocket to remain unchanged. Creating a probe with a knock out mutation would allow for a control protein with no binding to be used as a comparison to the wild type LecB and other mutants that exhibit binding. Wild type LecB contains only one lysine residue in its amino acid composition. Standard chemical methods to biotinylated proteins use the NH^{3+} side group of lysine to chemically add biotin labels to proteins. Once biotinylated, streptavidin linked dyes may be used as fluorophores with streptavidin and biotin being one of the strongest non-covalent bonds known in nature. All of the commercially available eukaryotic lectins purchased were biotinylated and engineering a protein with added lysine would facilitate similar biotinylation of the recombinant prokaryotic LecB being produced. Figure 3.4 shows the five important amino acid structures used in the mutagenesis of LecB.

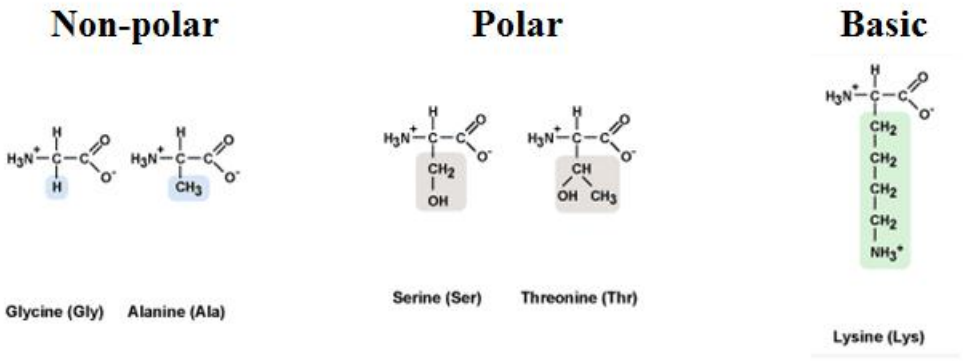


Figure 3.4: Important amino acid structures for site directed mutagenesis. Glycine and alanine are non-polar amino acids, serine and threonine are polar amino acids while lysine is a basic amino acid.

Primer Name	Sequence (5'-3')	Tm (°C)
LecB S24T for.	GCCAACTCGACCGGAACCCAGACGGTGAAC	68.5
LecB S24T rev.	GAAGGCGGTGACGCCGAACCG	62.2
LecB 3K for.	AAAAAGAAATCCATGGCAACACAAGGAGTGTTCACC	63.3
LecB 3K rev.	TCCGTGATGGTGATGGTGATG	54.4
LecB 3A for.	GCCAACGCGGCCGCAACCCAGACGGTGAACGTG	73.1
LecB 3A rev.	GAAGGCGGTGACGCCGAACCG	62.2

Table 3.1: Primer sequences for creating library of LecB mutants. Mutations are highlighted in blue. Tm is the melting temperature of the primer.

In table 3.1 the primer sequences to introduce the three LecB mutations described above are shown. The regions highlighted represent the changes that are to be implemented into the plasmid (see Section 2.11). All of the mutations were added using the forward primer. The PCR reaction was set up as described in section 2.12. The melting temperature for each primer is different as can be seen in table 3.1 which leads to varying annealing temperatures when setting up a PCR reaction. A multi-annealing temperature PCR was set up using a thermocycler that allowed for a maximum of 6 different annealing temperatures to be used. Originally three temperatures (55, 60 and 65 °C) were used and the optimal temperature could be optimised using these temperatures. In figure 3.5 the PCR products for LecB with added lysine are run on a 0.7% agarose gel. There is product formed from 55 to 65 °C annealing temperatures but the most product was found in the 60 °C reaction, analysed by nanodrop (NanoDrop ND-1000). The primers used have a phosphorylated 5'- end. This allows for the methylated template DNA to be digested using the restriction enzyme *Dpn*I (see section 2.13). *Dpn*I has a proclivity for methylated DNA and digests the template DNA to only leave the phosphorylated PCR product that now contains a mutated region. This PCR product is cleaned (see section 2.9) and the linear plasmid is ligated to re-circularise the DNA (see section 2.14).

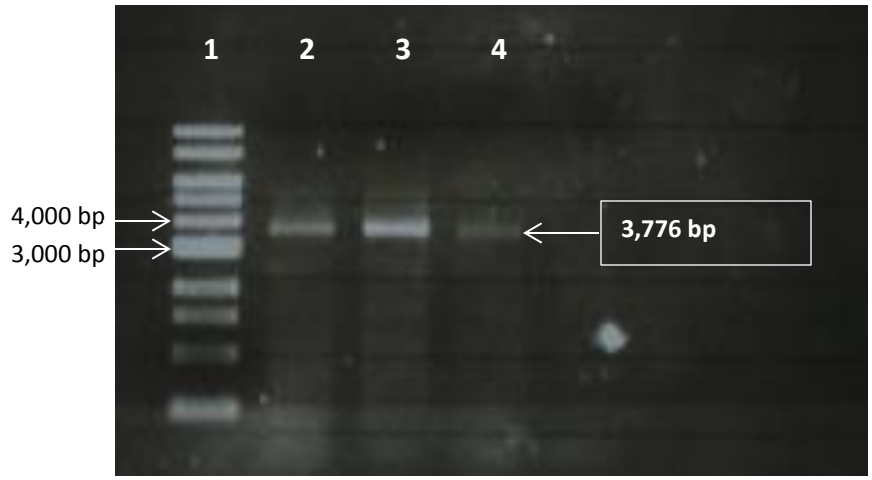


Figure 3.5: LecB PCR products for LecB 3K amplification. Lanes; 1) 1 kb DNA ladder, 2) PCR product annealing temperature 55 °C, 3) PCR product annealing temperature 60 °C, 4) PCR product annealing temperature 65 °C. Products run on a 0.7% agarose gel. Linear plasmids run just below 4,000 bp.

The ligated plasmid was transformed into competent JM109 *E. coli* cells (see Section 2.15) and these cells now contain an intact plasmid containing the LecB gene with the various mutations introduced. These plasmids were sequenced to ensure the mutations have taken place correctly. These plasmids are shown in figure 3.2. All of the plasmids sequenced were compared to the original LecB plasmid sequence. Figure 3.6 shows the three mutated LecB sequences compared to the reference LecB sequence. The amino acid sequences were obtained using the nucleotide sequence reads (see Section 2.10). The sequences were input into the online translate tool ExPASy (see section 2.29) and the amino acid sequences were aligned to show the incorportation of the amino acid differences using Multalin.

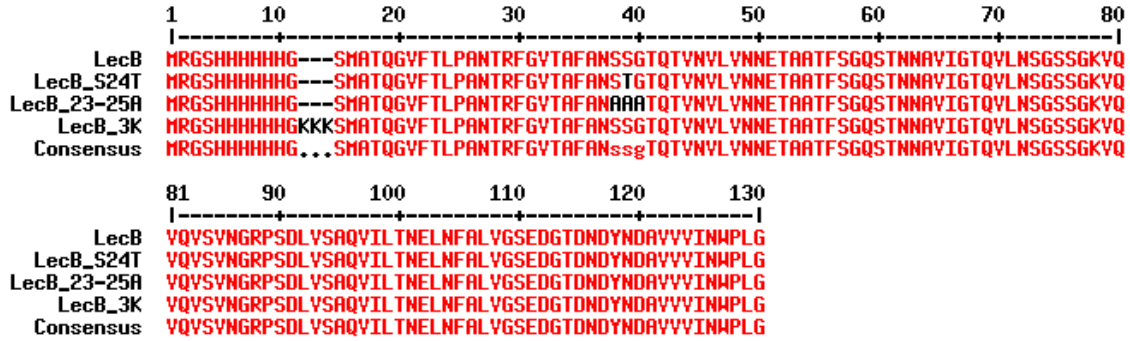


Figure 3.6: Amino acid sequence alignment of LecB and LecB mutants. LecB, LecB S24T, LecB 3A, LecB 3K were all aligned using the online tool multalin. Consensus sequences are in red while the black regions highlight the mutations.

3.3 Generation of eGFP-LecB Fusion Protein

Along with the three LecB mutants generated in section 3.2, another hypothesis was to engineer a LecB variant that could be easily fluorescently detected. This was achieved by fusing LecB with a fluorescent fusion protein partner. In this case the protein fused to the LecB was enhanced green fluorescent protein (eGFP). GFP is a fluorescent protein discovered in the jellyfish *Aequorea Victoria* (Zimmer 2002). Specific mutations to the amino acid sequence of GFP have substantially increased the fluorescence to generate eGFP. Through introducing a fluorescent fusion partner the necessity for biotinylation of LecB is removed and this protein probe can be used as is. GFP has been used in a wide variety of ways as a fusion partner because it does not interfere with the primary proteins function and also in the case of non-stable proteins it can increase its stability.

The LecB plasmid sequence contains the restriction site *BamH*I between the poly-histidine tag and start codon of the gene. *BamH*I recognises the nucleotide sequence GGATCC. The restriction enzyme cuts both strands of DNA leaving 5' prime overhangs. The LecB plasmid also contains a *Hind*III restriction site after the stop codon of the gene. *Hind*III recognises the nucleotide sequence AAGCTT. The LecB sequence was designed to have the GGATCC region in between the poly-histidine tag and the start codon of the gene sequence to allow for the gene to be excised from the plasmid. The sequence was also designed to have the AAGCTT region after the stop codon to allow the stop codon be intact when the gene is transferred into another plasmid. Using both of these restriction enzymes the LecB gene was successfully excised from the plasmid without its poly-histidine tag and with its stop codon intact. Figure 3.7 shows the LecB nucleotide sequence and also highlights where the restriction sites occur.

The eGFP gene sequence was modified to allow for it to be a fusion partner. The stop codon was removed and replaced with the GGATCC nucleotide sequence to allow for *BamH*I restriction. A short number of nucleotide bases after this GGATCC sequence there is an AAGCTT sequence that allows for *Hind*III restriction. Using the restriction enzymes *BamH*I and *Hind*III causes 5' overhang nucleotide sequences which facilitates ligation of the LecB gene into the cloning site of eGFP as well as keeping the inserted gene in the correct open reading frame. Sticky end ligations are more desirable than blunt end ligations. The eGFP gene was added to a pQE30 vector to allow for an *N*-terminal poly-histidine tag. Post poly-histidine tag the amino acid sequence still contained the 'GS' residues before the genes methionine start codon. To ensure the gene does not get cut at this region the nucleotide sequence is changed to GGATCT so *BamH*I will not cut at this region.

The eGFP-LecB fusion protein created retains the *BamH*I site between the two proteins, as well as the *Hind*III site post stop codon. These restriction sites allow for the gene to be removed and replaced with any *BamH*I and *Hind*III ended gene to allow for eGFP to be used as a fusion partner for other desired protein. Other projects running in unison with this project used this eGFP gene to create a similar fusion protein. The nucleotide sequences in figure 3.7, 3.8 and 3.9 highlight the recognition sites for the restriction enzymes used.

LecB wild type nucleotide sequence

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC ATG GCA ACA CAA
GGA GTG TTC ACC CTT CCC GCC AAC ACC CGG TTC GGC GTC ACC GCC TTC
GCC AAC TCG TCC GGA ACC CAG ACG GTG AAC GTG CTG GTC AAC AAC GAG
ACG GCC GCG ACC TTC AGC GGG CAA AGC ACC AAT AAC GCC GTC ATC GGC
ACC CAG GTG CTC AAC TCC GGC AGC AGT GGC AAG GTA CAG GTC CAG GTC
AGC GTC AAC GGC CGC CCC TCG GAT CTG GTC TCG GCA CAG GTA ATC CTG
ACC AAC GAG CTG AAC TTC GCC CTG GTC GGC TCT GAA GAC GGC ACC GAC
AAC GAC TAC AAC GAC GCC GTC GTG GTG ATC AAC TGG CCG CTC GGC TAG
TAA AAG CTT AAT

Figure 3.7: DNA nucleotide sequence of LecB gene. The sequence shows four codons past the stop codon to highlight the *Hind*III region. Red = start codon, Yellow = stop codon, Green = *BamH*I recognition site, Blue = *Hind*III recognition site.

Modified eGFP-nucleotide sequence

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCT ATG AGT AAA GGA
GAA GAA CTT TTC ACT GGA GTT GTC CCA ATT CTT GTT GAA TTA GAT GGT
GAT GTT AAT GGG CAC AAA TTT TCT GTC AGT GGA GAG GGT GAA GGT GAT
GCA ACA TAC GGA AAA CTT ACC CTT AAA TTT ATT TGC ACT ACT GGA AAA CTA
CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT CTC ACT TAT GGT GTT CAA
TGC TTT GCG AGA TAC CCA GAT CAT ATG AAA CAG CAT GAC TTT TTC AAG
AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA AGA ACT ATA TTT TTC AAA
GAT GAC GGG AAC TAC AAG ACA CGT GCT GAA GTC AAG TTT GAA GGT GAT
ACC CTT GTT AAT AGA ATC GAG TTA AAA GGT ATT GAT TTT AAA GAA GAT
GGA AAC ATT CTT GGA CAC AAA TTG GAA TAC AAC TAT AAC TCA CAC AAT
GTA TAC ATC ATG GCA GAC AAA CAA AAG AAT GGA ATC AAA GTT AAC TTC
AAA ATT AGA CAC AAC ATT GAA GAT GGA AGC GTT CAA CTA GCA GAC CAT
TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT GTC CTT TTA CCA GAC
AAC CAT TAC CTG TCC ACA CAA TCT GCC CTT TCG AAA GAT CCC AAC GAA
AAG AGA GAC CAC ATG GTC CTT CTT GAG TTT GTA ACA GCT GCT GGG ATT
ACA CAT GGC ATG GAT GAA CTA TAC AAA GGA TCC GCA TGC GAG CTC GGT
ACC CCG GGT CGA CCT GCA GCC AAG CTT AAT TAG

Figure 3.8: DNA nucleotide sequence of modified eGFP gene. The modified eGFP gene contains the added restriction sites necessary to create a fusion protein. Red = start codon, Yellow = stop codon, Green = *BamH*I recognition site, Blue = *Hind*III recognition site, Purple = *Acc*I recognition site.

eGFP-LecB nucleotide sequence

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCT ATG AGT AAA GGA
GAA GAA CTT TTC ACT GGA GTT GTC CCA ATT CTT GTT GAA TTA GAT GGT
GAT GTT AAT GGG CAC AAA TTT TCT GTC AGT GGA GAG GGT GAA GGT GAT
GCA ACA TAC GGA AAA CTT ACC CTT AAA TTT ATT TGC ACT ACT GGA AAA CTA
CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT CTC ACT TAT GGT GTT CAA
TGC TTT GCG AGA TAC CCA GAT CAT ATG AAA CAG CAT GAC TTT TTC AAG
AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA AGA ACT ATA TTT TTC AAA
GAT GAC GGG AAC TAC AAG ACA CGT GCT GAA GTC AAG TTT GAA GGT GAT
ACC CTT GTT AAT AGA ATC GAG TTA AAA GGT ATT GAT TTT AAA GAA GAT
GGA AAC ATT CTT GGA CAC AAA TTG GAA TAC AAC TAT AAC TCA CAC AAT
GTA TAC ATC ATG GCA GAC AAA CAA AAG AAT GGA ATC AAA GTT AAC TTC
AAA ATT AGA CAC AAC ATT GAA GAT GGA AGC GTT CAA CTA GCA GAC CAT
TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT GTC CTT TTA CCA GAC
AAC CAT TAC CTG TCC ACA CAA TCT GCC CTT TCG AAA GAT CCC AAC GAA
AAG AGA GAC CAC ATG GTC CTT CTT GAG TTT GTA ACA GCT GCT GGG ATT
ACA CAT GGC ATG GAT GAA CTA TAC AAA GGA TCC ATG GCA ACA CAA GGA
GTG TTC ACC CTT CCC GCC AAC ACC CGG TTC GGC GTC ACC GCC TTC GCC
AAC TCG TCC GGA ACC CAG ACG GTG AAC GTG CTG GTC AAC AAC GAG ACG
GCC GCG ACC TTC AGC GGG CAA AGC ACC AAT AAC GCC GTC ATC GGC ACC
CAG GTG CTC AAC TCC GGC AGC AGT GGC AAG GTA CAG GTC CAG GTC AGC
GTC AAC GGC CGC CCC TCG GAT CTG GTC TCG GCA CAG GTA ATC CTG ACC
AAC GAG CTG AAC TTC GCC CTG GTC GGC TCT GAA GAC GGC ACC GAC AAC
GAC TAC AAC GAC GCC GTC GTG GTG ATC AAC TGG CCG CTC GGC TAG TAA
AAG CTT AAT

Figure 3.9: DNA nucleotide sequence of fusion gene eGFP-LecB. The fusion protein still contains the *BamH*I and *Hind*III restriction sites to allow for removal of LecB gene. Red = start codon, Yellow = stop codon, Green = *BamH*I recognition site, Blue = *Hind*III recognition site.

In between the *BamH*I and *Hind*III recognition sites on the modified eGFP sequence there are a number of other restriction sites. One of interest is GTCGAC which is a recognition site for the restriction enzyme *Acc*I. The pQE30 plasmid containing the eGFP gene contains 3 *Acc*I recognition sites but when the plasmid is cut with *BamH*I and *Hind*III one of the sites is removed. The LecB gene is then ligated into the region between *BamH*I and *Hind*III resulting in the pQE30-eGFP plasmid containing three *Acc*I recognition sites, while the pQE30-eGFP-LecB plasmid contains only two. To select viable clones from the transformation plate single colonies were streaked (see section 2.6). The single colonies then found on the streak plate were used to inoculate overnight cultures and the plasmids from these cultures were purified (see sections 2.7). Before the samples were sent for sequencing, a restriction digest using *Acc*I was carried out to confirm the presence of the LecB gene in the newly formed plasmid. In figure 3.12 the agarose gel shows the two undigested eGFP-LecB plasmids selected along with the control pQE30-eGFP plasmid. The *Acc*I restriction carried out on the eGFP only plasmid will create three bands at sizes 2,712, 1,174 and 313 while the *Acc*I restriction carried out on the eGFP-LecB plasmid will create two bands at sizes 2,712 and 1,802. Figure 3.13 shows the 3 plasmids restricted, these are pQE30 plasmids containing either the eGFP-LecB fusion gene or just the eGFP gene. The smaller band of 313 base pairs did not appear but the smaller band of 1,174 compared to 1,802 base pairs can be visualised. Both of the plasmids selected indicate that the LecB gene had successfully been ligated into the gene. Figures 3.10 and 3.11 show whole plasmid images of pQE30 with the eGFP gene and the eGFP-LecB gene and where these restriction sites are as well as the other desirable traits of the plasmid for protein expression, namely the poly-histidine tag, ampicillin resistance gene and promoter region.

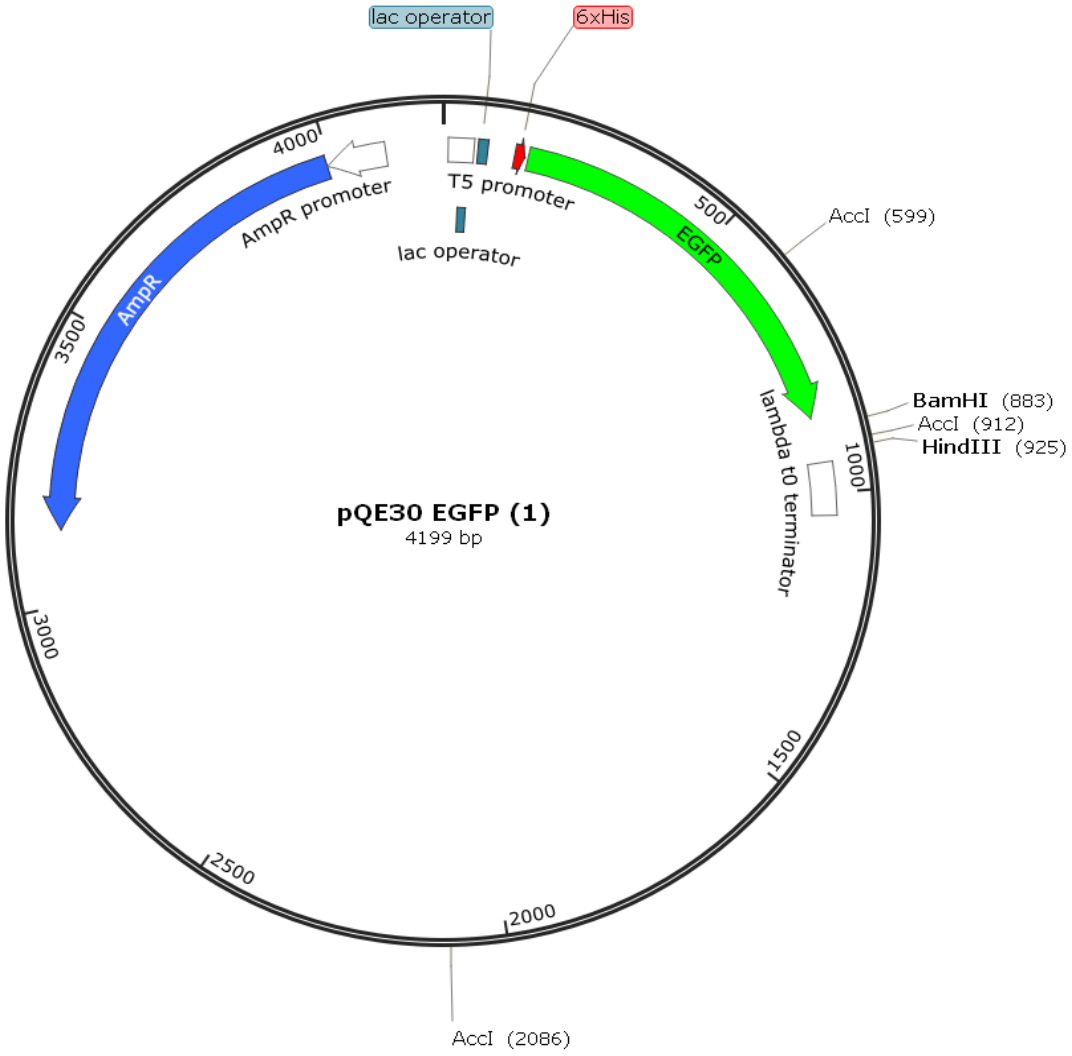


Figure 3.10: pQE30 plasmid containing the eGFP gene with restriction sites *BamH*I, *Hind*III and *Acc*I highlighted. The plasmid shows the size of the plasmid as well as the lac operator, T5 promoter and Ampicillin resistance gene. The gene map was generated using Snapgene (see Section 2.29) and only some of the important features of the plasmid are highlighted.

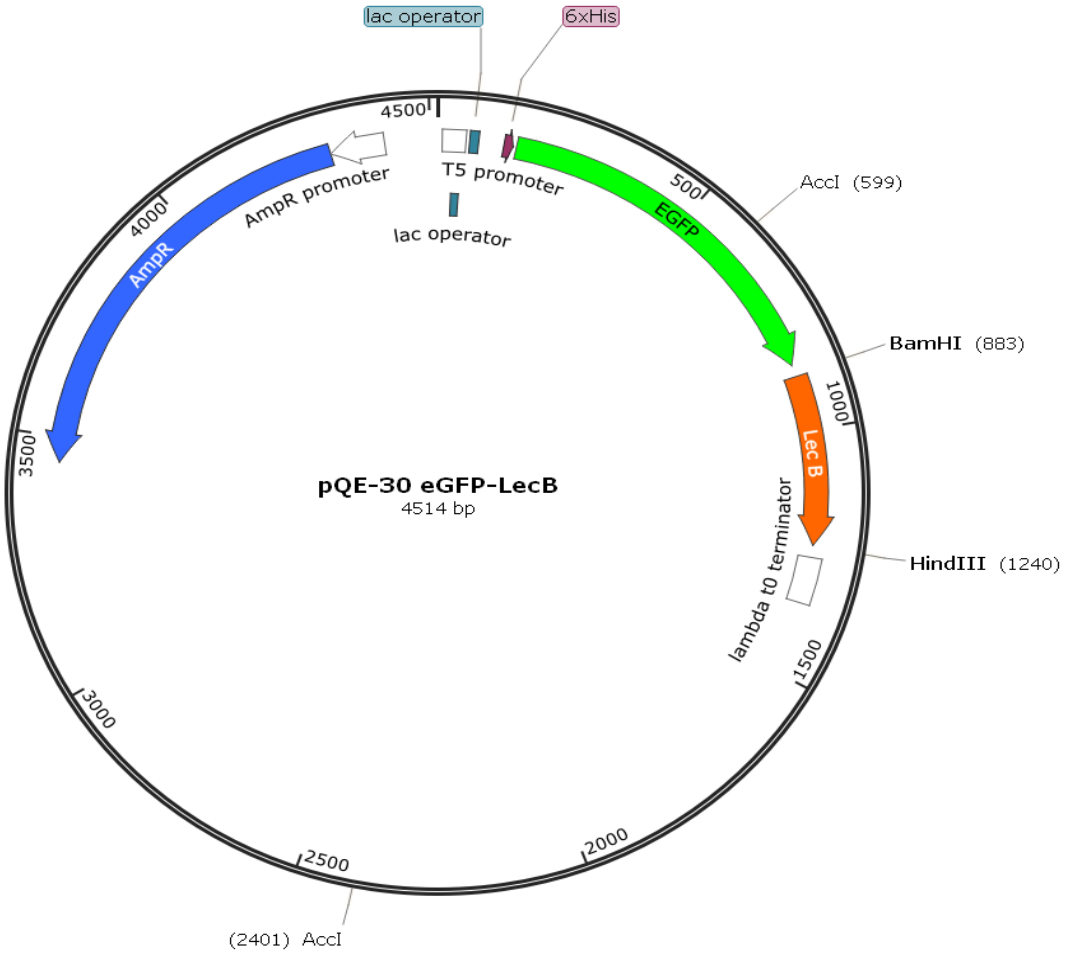


Figure 3.11: pQE30 plasmid containing the fusion protein eGFP-LecB with restriction sites *BamH*I, *Hind*III and *Acc*I highlighted. The plasmid shows the size of the plasmid as well as the lac operator, T5 promoter and Ampicillin resistance gene. The gene map was generated using Snapgene (see Section 2.29) and only some of the important features of the plasmid are highlighted.

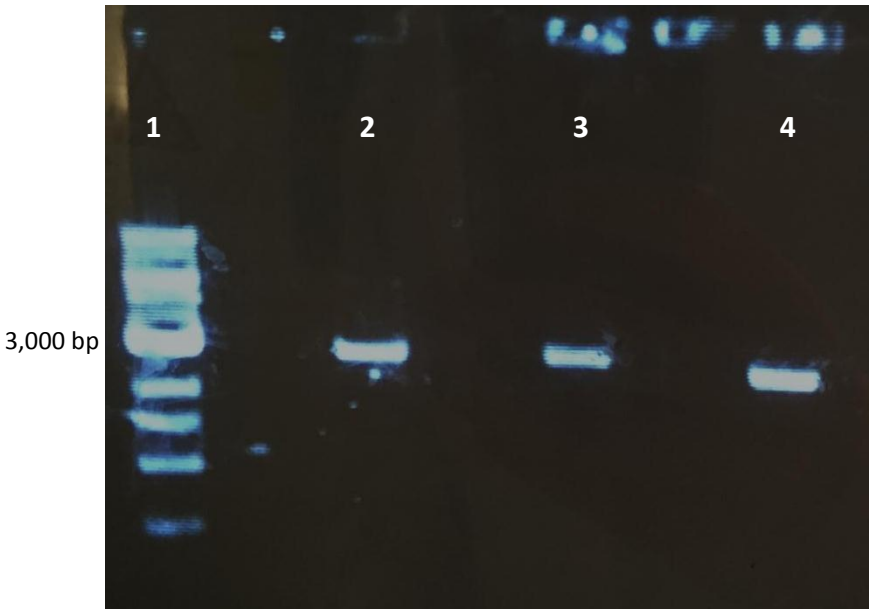


Figure 3.12: 1% (w/v) agarose gel containing uncut supercoiled plasmids. Lanes; 1) 1 kb DNA ladder, 2) eGFP-LecB plasmid clone 1, 3) eGFP-LecB plasmid clone 2, 4) eGFP plasmid. Gel indicates intact purified plasmid.

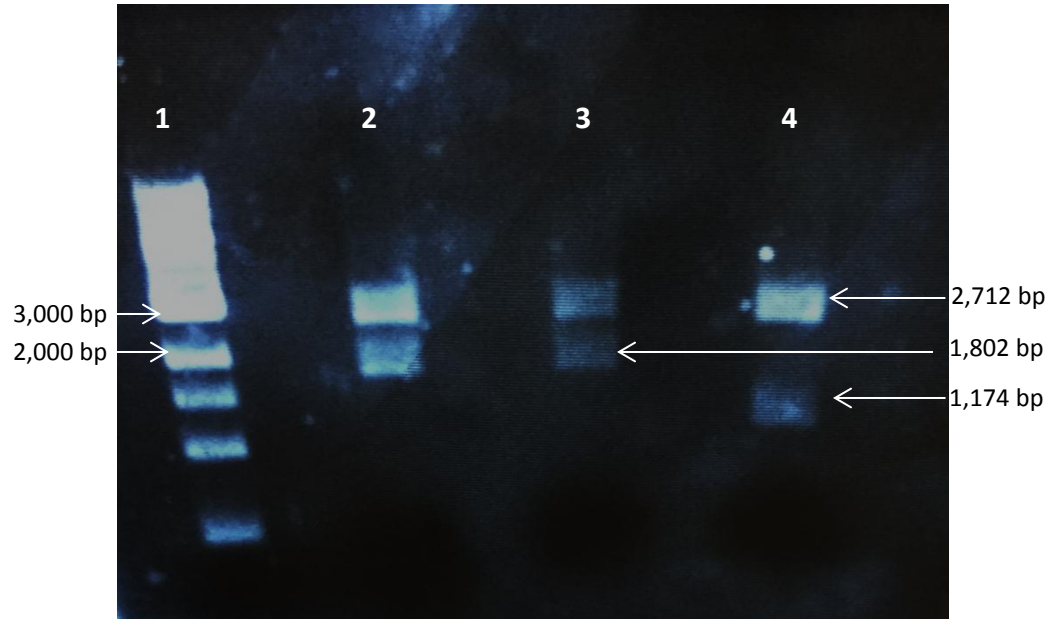


Figure 3.13: 1% (w/v) agarose gel containing plasmids cut with the restriction enzyme AccI. Lanes; 1) 1 kb DNA ladder, 2) eGFP-LecB plasmid clone 1, 3) eGFP-LecB plasmid clone 2, 4) eGFP plasmid.

The restriction digest confirmed that LecB had successfully been ligated into the pQE30-eGFP plasmid. Both plasmids were sent for DNA sequencing and both clones contained the correct plasmid in the correct reading frame. The amino acid sequences of eGFP, LecB and eGFP-LecB were run through alignment software to show that the sequences were correct. This can be seen in figures 3.14 and 3.15. LecB was introduced to the eGFP plasmid and not the other way around to ensure the poly-histidine tag was still accessible for IMAC purification. As stated previously the *C*- terminus of the LecB protein has a role in the proteins ability to bind its sugar and any changes to this region can have implications for the proteins ability to remain active. There are only 2 amino acids, the BamHI restriction site 'GS', between the LecB and eGFP, often there would be a longer linker molecule containing a stretch of amino acids that would create separation between both of the proteins. The BamHI restriction site allows for the option of addition amino acids to be added generating this linker molecule. Originally a protein fusion without a linker was hypothesised and a linker would be added if the protein lost its ability to bind its targets or there was a drop in protein binding affinity. The purified eGFP-LecB protein remained active and an amino acid linker was not necessary.

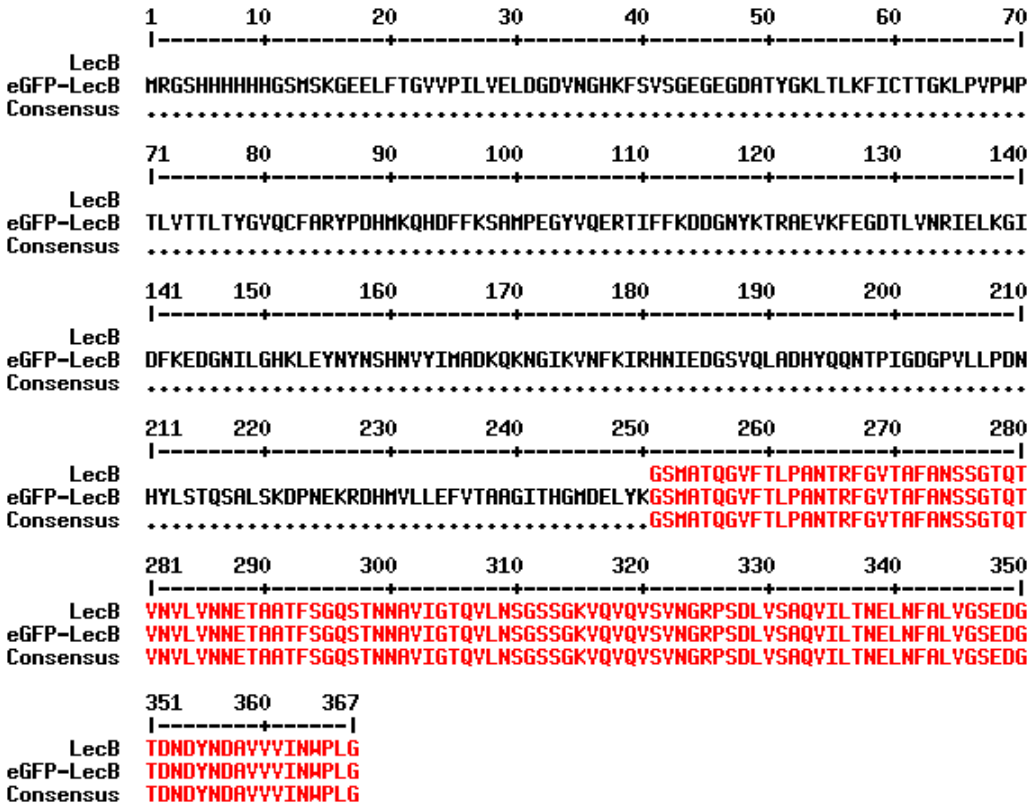


Figure 3.14: LecB and eGFP-LecB amino acid alignment. EGFP-LecB is aligned against the reference LecB amino acid sequence to show sequence consensus.

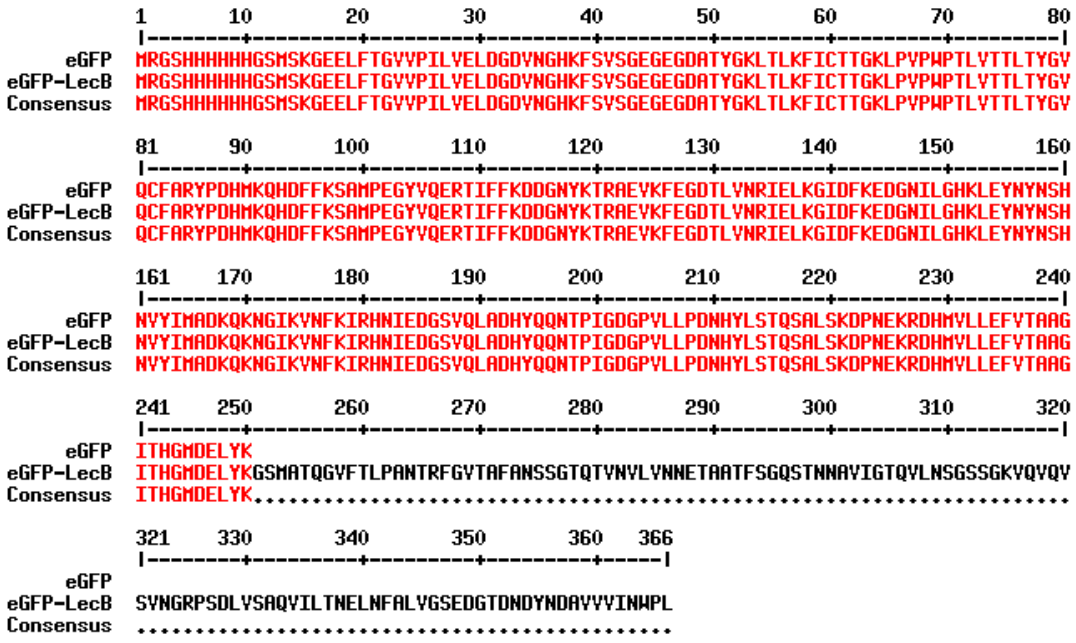


Figure 3.15: EGFP and eGFP-LecB amino acid sequence alignment. EGFP-LecB is aligned against the reference eGFP amino acid sequence to show sequence consensus.

The library of prokaryotic LecB variants now in totality contained five different proteins. They are; wild type LecB, LecB S24T, LecB 3A, LecB 3K and eGFP-LecB. Once these sequences were confirmed through DNA sequencing and sequence alignment the *E. coli* stocks containing these proteins were cultured and two stocks were created, one working stock at -20 °C and one long term storage stock at -80 °C, for each protein. All nucleotide and amino acid sequence data is available in appendix A.2.

3.4 Expression and Purification of LecB and LecB variants

After the sequences were confirmed the purified plasmid was transformed into competent JM109 *E. coli* or competent KRX *E. coli* for eGFP-LecB. From these transformation spread plates colonies were selected and re-streaked for single colonies (see Section 2.6). From these streak plates overnight cultures to test for expression were grown up (see Section 2.16). From these samples the clone that produced the most desired protein was kept for long term storage, again a working stock at -20 °C and a long term stock at -80 °C. In figure 3.16 the expression checks for different LecB S24T and LecB 3A clones are analysed for protein production.

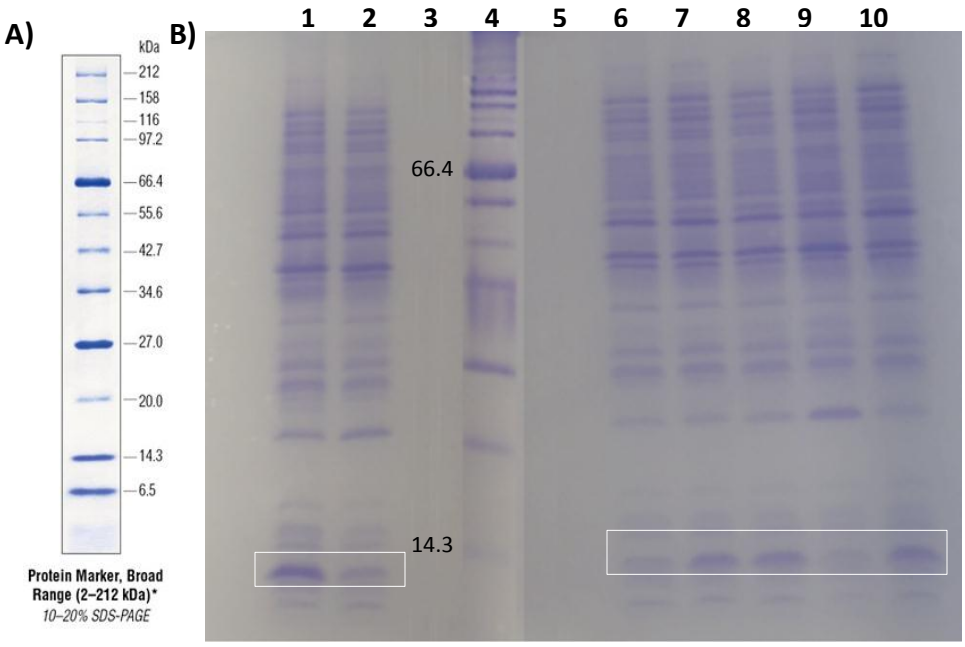


Figure 3.16: Expression analysis of different LecB S24T and LecB 3A clones on SDS-PAGE. A) Reference broad range protein ladder. B) Lanes; 1-2) LecB S23T clones 1 and 2 respectively, 3) blank, 4) Broad range protein ladder, 5) blank, 6-10) LecB 3A clones 1-5 respectively. The white boxes highlight the protein of interest in each sample. 66.4 and 14.3 are the kDa marker size in the ladder at those positions.

Expression analysis was carried out on a large number of clones to select a specific clone that expressed high levels of the desired protein. Five clones containing the library of five LecB variants were selected and stored for both short and long term use. Once a clone is selected and confirmed to have a high level of protein expression and the correct plasmid with the desired mutations present it was the only clone used for future expression and purification. All other clones were discarded leaving the five clones with our five LecB variants as the only clones in short and long term storage.

These *E. coli* stocks were expressed for large scale protein production and purification. Following the methods outlined in sections 2.16 and 2.17, cells were grown up to a large quantity and the protein was induced for overexpression during exponential bacterial cell growth. The cells are then lysed to release the cleared lysate containing our poly-histidine tagged protein. The cleared lysate was run through a nickel IMAC column to selectively purify all of the desired protein (see Section 2.18). The column is subjected to various wash steps containing different levels of free imidazole and finally a high imidazole elution step to remove the bound protein. The elution fractions are collected in 1 mL aliquots and run on an SDS-PAGE gel along with the cleared lysate, unbound flow through and wash steps to visualise the purity of the protein and also identify the fractions containing the most protein. The nickel resin had a maximum saturation point for allowing poly-his_6 tagged proteins. If this saturation point is reached, the flow through of unbound protein will show our desired protein. This can be seen in figure 3.17, which is a purification of wild type LecB, and figure 3.18, which is a purification of LecB S24T. The LecB protein reached saturation as did the LecB S24T to a lesser degree. In appendix A.1 there are images of earlier purification SDS-PAGE gels that contain contaminating bands in the elution fractions or that show issues encountered when attempting to purify each LecB variant protein.

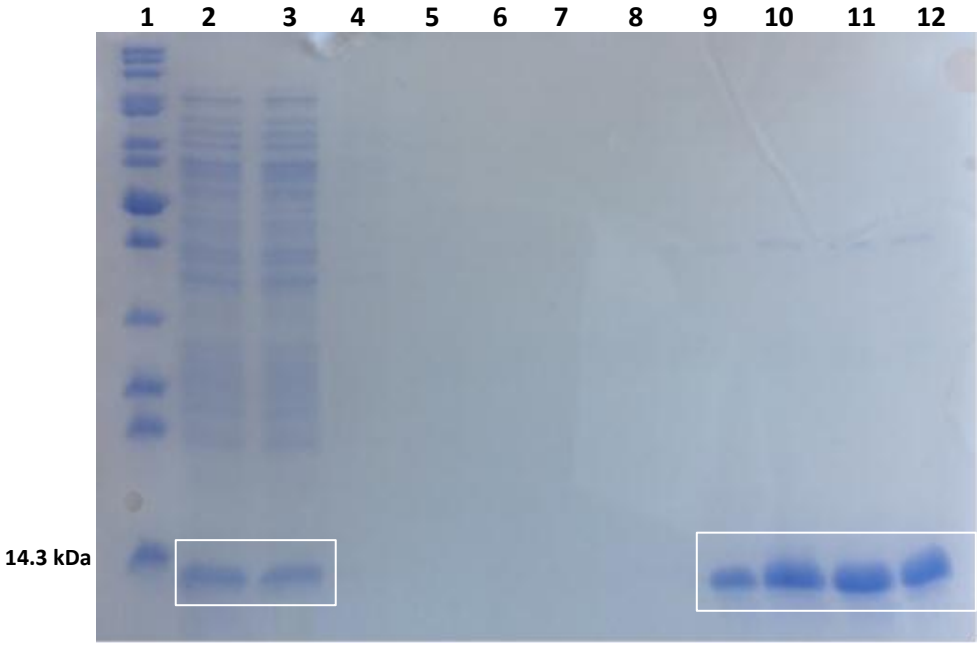


Figure 3.17: Purification of *N*-terminally poly-his$_6$ tagged LecB from cleared cell lysate using Ni-NTA resin. Analysis of fractions collected throughout purification of LecB on 15% (v/v) SDS-PAGE gel. Lanes; 1) Broad range protein ladder, 2) Cleared filtered lysate, 3) Unbound flow through, 4) Lysis buffer containing 20 mM imidazole wash, 5) Lysis buffer containing 50 mM imidazole wash, 6) Lysis buffer containing 80 mM imidazole wash, 7) Lysis buffer containing 100 mM imidazole wash, 8) Elution fraction 1, 9-12) Elution fractions 2-5 respectively. 14.3 is the kDa marker size in the ladder at that position. The white box signifies the LecB protein.

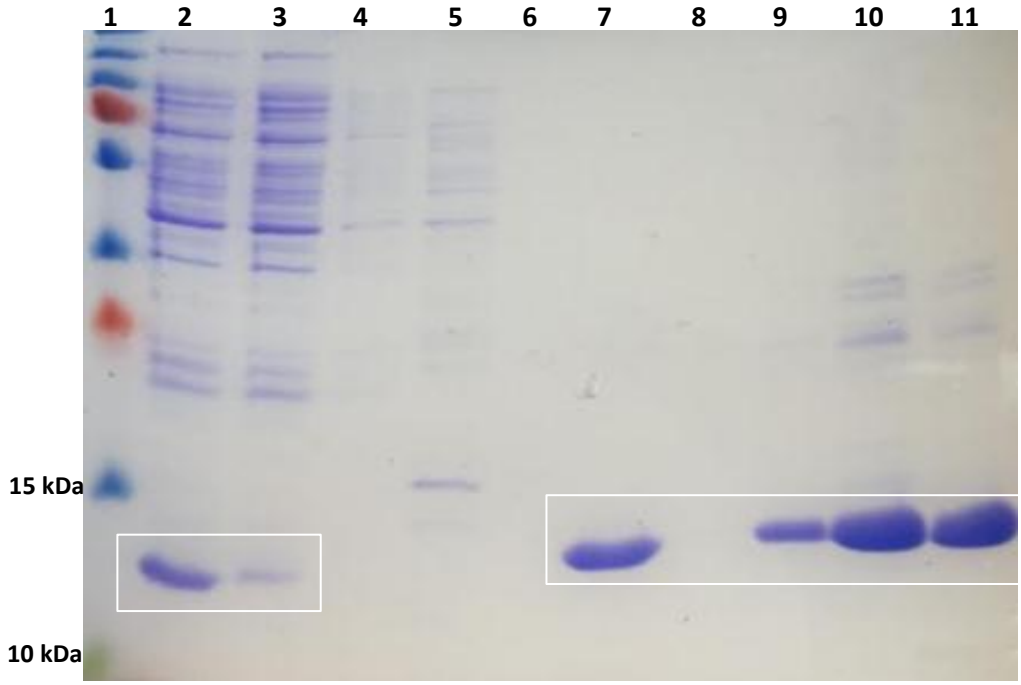


Figure 3.18: Purification of *N*-terminally poly-his$_6$ tagged LecB S24T from cleared cell lysate using Ni-NTA resin. Analysis of fractions collected throughout purification of LecB S24T on 15% SDS-PAGE gel. Lanes; 1) Three colour Broad range protein ladder, 2) Cleared filtered lysate, 3) Unbound flow through, 4) Lysis buffer containing 20 mM imidazole wash, 5) Lysis buffer containing 50 mM imidazole wash, 6) Lysis buffer containing 80 mM imidazole wash, 7) Elution fraction 1, 8) Problem with the well no sample added, 9-11) Elution fractions 2-5 respectively. 15 and 10 are the kDa marker sizes in that ladder at that position. The white box signifies the LecB S24T protein.

During the original purification with LecB there were purifications that contained some contaminating bands and for this reason there were additional wash steps with increasing levels of imidazole. These increasing concentrations of imidazole were to wash away any higher proteins that were possibly weakly bound to the nickel resin based on surface histidine residues. In figure 3.17 there are LecB bands in both the cleared filtered lysate and the unbound flow through; this represents the LecB saturating the nickel resin. The elution fractions contain a high level of purified LecB. Both figures 3.18 and 3.19 show that the majority of the poly-his_6 tagged protein was retained by the resin and very little was detected in the unbound flow through. In appendix A.1 other SDS-PAGE gels will also confirm this.

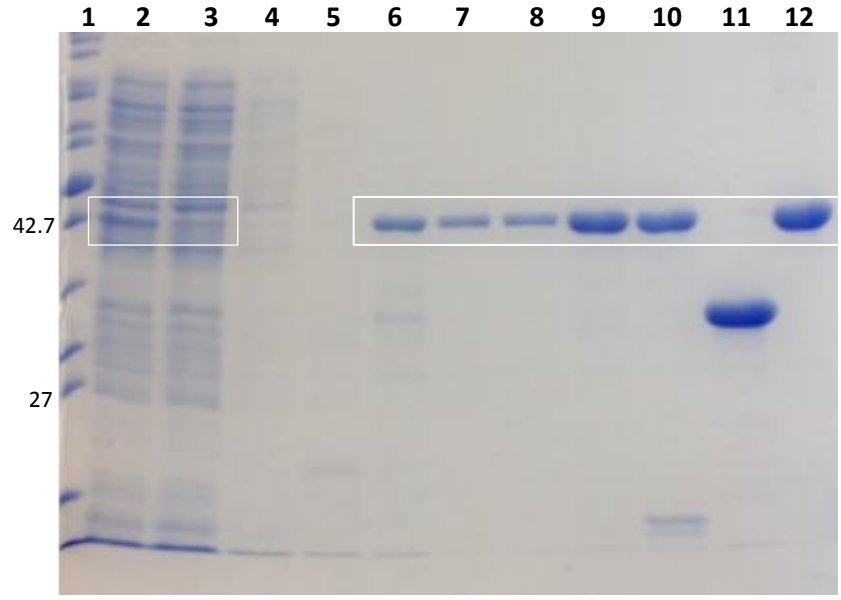


Figure 3.19: Purification of *N*-terminally poly-his$_6$ tagged eGFP-LecB from cleared cell lysate using Ni-NTA resin. Analysis of fractions collected throughout purification of eGFP-LecB on 15% (v/v) SDS-PAGE gel. Lanes; 1) Broad range protein ladder, 2) Cleared filtered lysate, 3) Unbound lysate, 4) Lysis buffer containing 20 mM imidazole wash, 5) Lysis buffer containing 80 mM imidazole wash, 6) Elution fraction 1, 7) Elution fraction 2, 8) Elution fraction 3, 9) Elution fractions 4-7 pooled, 10) Elution fractions 8-9 pooled, 11) Purified eGFP, 12) Buffer exchanged eGFP-LecB from pooled elution fractions 4-7. The eGFP purified protein was used for another project in the lab and was used for size reference as well as the broad range protein ladder. 42.7 and 27 are the kDa marker size in the ladder at those positions. The white box signifies the eGFP-LecB fusion protein.

Based on other purifications only two wash steps (see Figure 3.19), of 20 mM imidazole and 80 mM imidazole were used before elution. In figure 3.19 the eGFP-LecB band can be seen in the cleared filtered lysate but not in the unbound flow through. This shows the protein did not exceed the saturation point of the nickel resin and the maximum amount of protein was retained and purified. Both a broad range protein marker and purified eGFP were used as molecular size markers to compare eGFP-LecB.

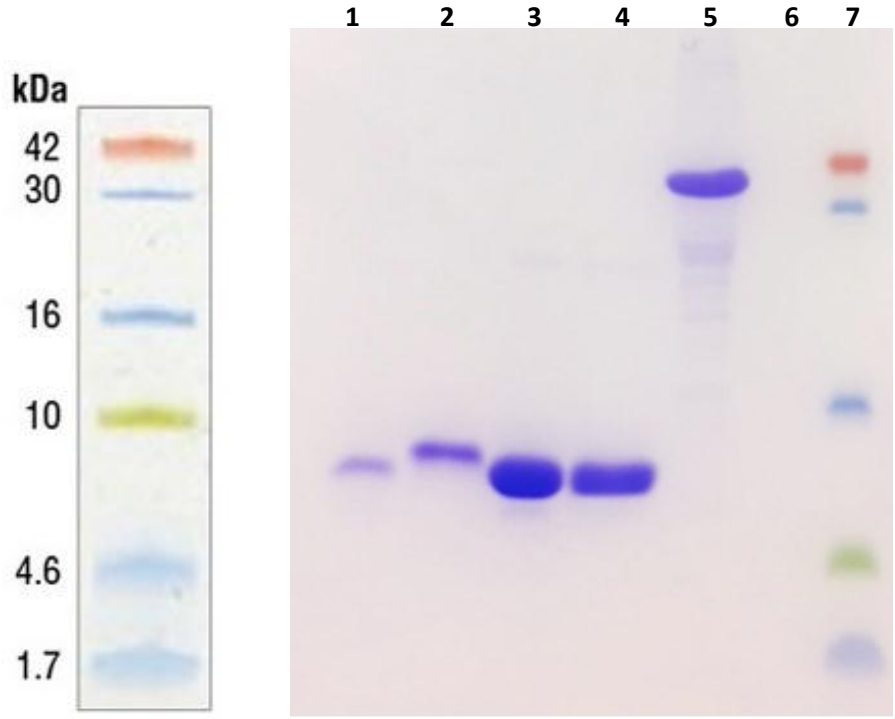


Figure 3.20: Purified fractions of the five LecB variants on a 15% SDS-PAGE gel using a low weight molecular marker. These are the five LecB variant protein samples that were purified, buffer exchanged and characterised run on a 15% SDS-PAGE gel to show purity. Lanes; 1) LecB, 2) LecB 3K, 3) LecB S24T, 4) LecB 3A, 5) eGFP-LecB, 6) Blank, 7) Low weight molecular marker. The Low molecular weight marker is from Cell Signalling Technologies®.

All the proteins were expressed and purified in similar ways. The only difference between any purification is the number of wash steps carried out with differing levels of imidazole. All protein elutions were carried out using 250 mM imidazole and all proteins were buffer exchanged and stored in PBS (see section 2.19). Figure 3.20 shows the library of purified LecB variants created using site directed mutagenesis. These proteins showed a high level of purity on SDS-PAGE gels.

3.5 Quantitation of purified LecB proteins

Proteins can be quantified in a number of ways. In this project all proteins were typically quantified using the bicinchoninic acid (BCA) assay and supplemented by spectrometer absorbance readings at 280 nm (Section 2.21). To quantify proteins using a BCA assay, a standard curve was generated using bovine serum albumin (BSA) standards. The standard curve was created using a working range of 0-2000 µg/mL that created a linear trend line once the data points are plotted. This generates an equation of the line that was used to quantify protein concentrations of LecB samples. The concentration of the sample protein was calculated based on the standard curve.

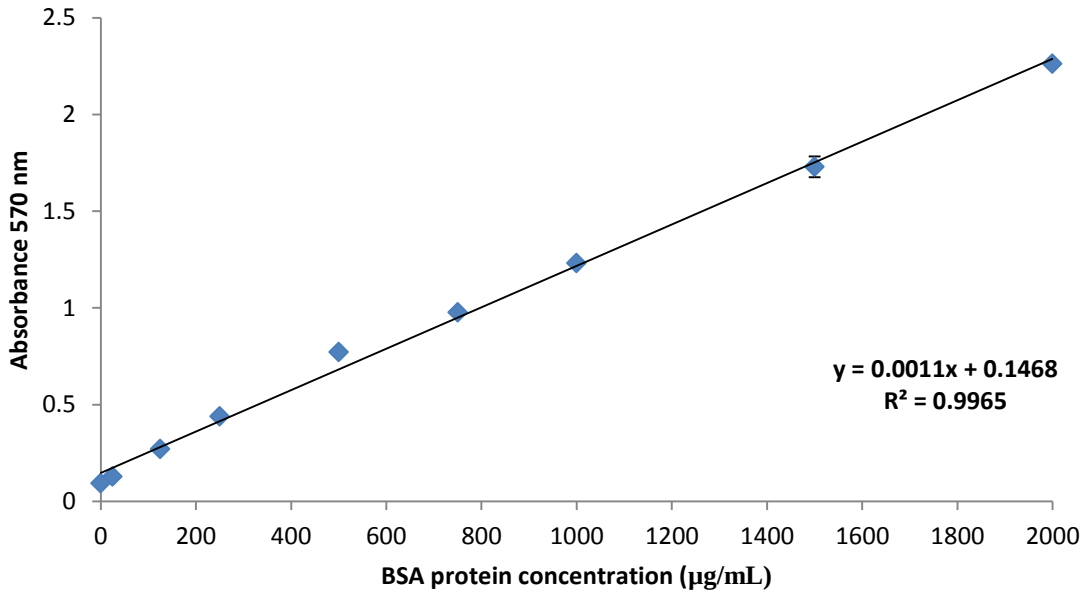


Figure 3.21: BCA standard curve using BSA standards. BSA standards ranging from 0-2000 µg/mL were using to create a linear trend line. The equation of the line and R^2 value are displayed on the chart. The R^2 value is a statistical measure of how close the data fits a regression line.

Sample	Absorbance (570 nm)	Concentration (mg/mL)
LecB	4.023	3.52
LecB S24T	4.635	4.08
LecB S23A;S24A;G25A	0.988	0.77
eGFP-LecB	0.881	0.67
LecB 3K	1.882	1.58
*LecB 3K biotinylated	1.643	1.36
*LecB S24T biotinylated	1.502	1.23

Table 3.2: Protein concentrations for the purified LecB mutants derived from BCA assays. Absorbance readings based off triplicate analysis of each individual protein. Absorbance values have been multiplied by there dilution factors. *LecB 3K that has gone through the biotinylation process after using the concentration of LecB 3K.

All samples were tested in triplicate to increase the accuracy of the quantification. The samples were also tested as a neat sample as well as diluted 1:3, 1:5 and 1:10 to ensure the values fit within the range of standards used, i.e. if the neat sample contained above 2 mg/mL it would not fit in our working concentration range. The proteins at lower concentrations per mL were lower due to the fact there was a higher volume of buffer exchanged protein. After quantification samples are aliquot into working stocks at certain concentrations, e.g. 500-1000 µg/mL. Proteins at high concentrations can aggregate and fall out of solution so all proteins are stored between 500-1000 µg/mL. As an alternative method, spectrometer readings at a 260 nm wavelength helped to confirm the protein concentrations were accurate (see Section 2.21).

3.6 Glycan binding characterisation of LecB and LecB mutants

LecB is originally compared to other eukaryotic lectins to characterise its binding profile. LecB is known to bind fucose with a high affinity and mannose to a lesser affinity. It is then tested against mannose and fucose linked to BSA as well as other glycoproteins that are known to contain both mannose and fucose. The technique used to test for lectin binding is called the enzyme linked lectin assay (ELLA) which is described in section 2.23. This technique was used to test the activity of the various LecB mutants produced throughout the project as well as test the accuracy and activity of commercially purchased eukaryotic lectins. *Aleuria Aurantia* lectin (AAL) and *Ulex Europarus* agglutinin I (UEA-I) are two eukaryotic lectins that both bind fucose residues, although fucose residues with different linkages. AAL binds preferentially to fucose residues that are linked to GlcNAc through an α-1,6 linkage or fucose linked to N-acetyllactosamine with an α-1,3 linkage, whereas UEA-I preferentially binds to fucose residues linked through an α-1,2 linkage. LecB has shown to bind various L-fucose residues with a very high affinity, while also binding to D-mannose. The ELLA was first carried out using commercially available eukaryotic lectins with defined binding specificities. Commercial glycoproteins; asialofetuin, fetuin and bovine mucin, were used to analyse the binding of the commercial lectins, Con A, RCA and Jacalin. This analysis can be viewed in figure 3.22. The binding specificity of these three lectins is shown in section 2.24, along with all other commercially purchased eukaryotic lectins. These lectins were probed at 5 µg/mL and each of their individual competitive inhibiting sugars was used. The sugar concentrations used were at the concentrations shown to have the highest inhibitory effect on the lectins. Some eukaryotic lectins show a very high affinity for glycans in complex and are not fully inhibited by competitive free sugars. Jacalin and RCA are not fully inhibited by the addition of galactose but there is a significant decrease in binding observed. Con A is fully inhibited by a mix of α-methylmannoside and α-methylglucoside. The binding activity of each commercially purchased eukaryotic lectin was validated by ELLA analysis.

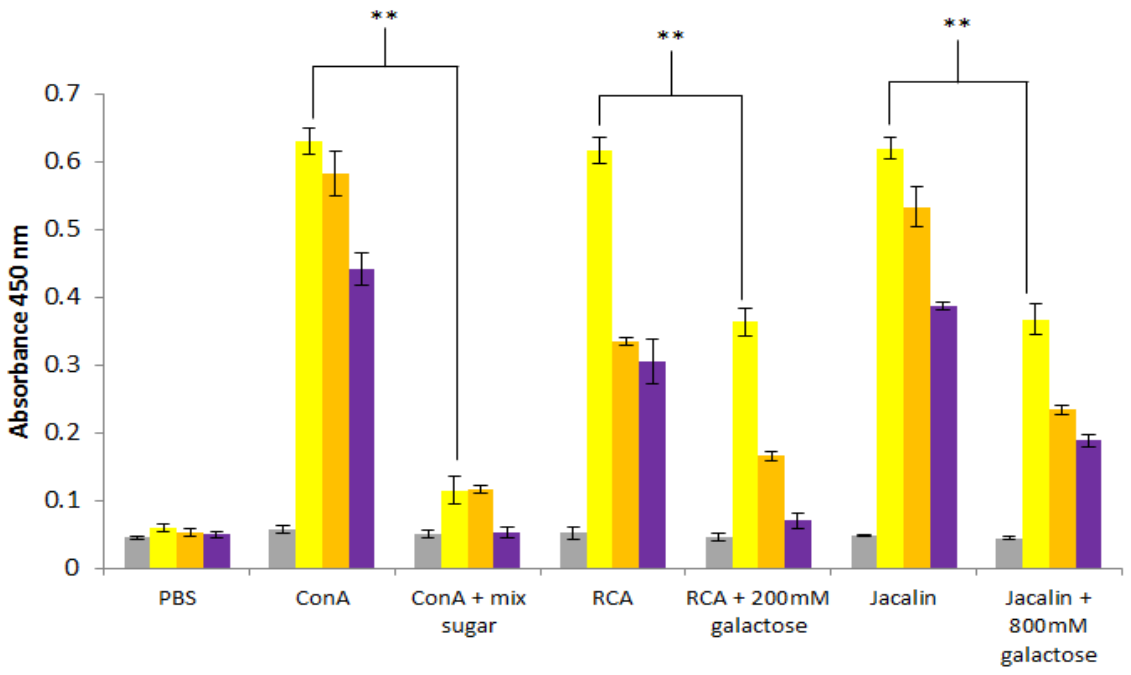


Figure 3.22: ELLA analysis of commercial eukaryotic lectins with their corresponding competitive inhibitory sugars. All lectins and glycoproteins were used at 5 µg/mL. The mixed sugar used for ConA inhibition is a 200 mM mix of α-methylmannoside and α-methylglucoside. Bars on the chart; Grey = PBS, Yellow = Asialofetuin, Orange = Fetuin, Purple = Bovine mucin. The galactose used for RCA and Jacalin competitive inhibition was D-galactose. Data are means ± standard deviations. ** represents $P < 0.001$, n=3.

The first comparison carried out for LecB was to compare its binding to AAL and UEA-I. Once purified, the recombinant lectins were used as controls to see if the anti-histidine antibody can successfully detect the protein. To do this the ELLA is partially changed with the lectin being allowed to bind to the plate and then blocking solution added. After this the proteins are probed with an anti-histidine antibdody to determine the available signal. In figure 3.23 the recombinant lectins LecB and LecB S24T confirm the antibodies ability to bind the His target. In figure 3.24 the three lectins, AAL, UEA-I and LecB, were tested for binding to three commercially purchased glycoproteins; Asialofetuin, Thyroglobulin and Bovine Mucin. Bovine mucin acts as a negative control

for LecB, while LecB binds to asialofetuin and thyroglobulin. There was also a competitive sugar inhibition carried out to highlight the binding specificity of the probe, the competitive sugar in the case of these three lectins was 100 mM L-fucose. LecB has a high affinity for thyroglobulin and a lesser affinity for asialofetuin. LecB doesn't appear to bind to bovine mucin while both AAL and UEA-I appear to bind to bovine mucin with varying levels of binding observed. All three lectins were completely inhibited using 100 mM L-fucose proving that the binding is specific through the glycan binding pocket and not any other protein-protein interaction.

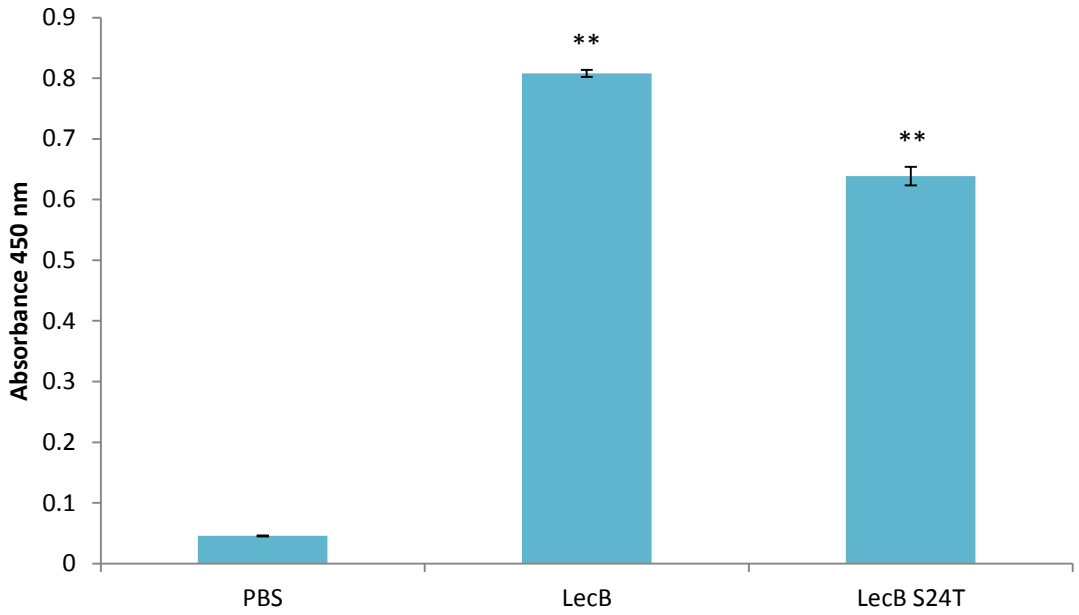


Figure 3.23: ELLA analysis to determine anti-poly-histidine antibody can detect LecB and LecB mutants. LecB and LecB S24T were added to the F96 MAXISORP NUNC 96-well plate to allow the protein to bind and then after a blocking step the proteins were probed with anti-poly-histidine antibody (monoclonal produced in mouse –H1029). Data are means ± standard deviations. ** represents $P < 0.001$, n=3. The average means of LecB and LecB S24T were tested against the average mean for the PBS control.

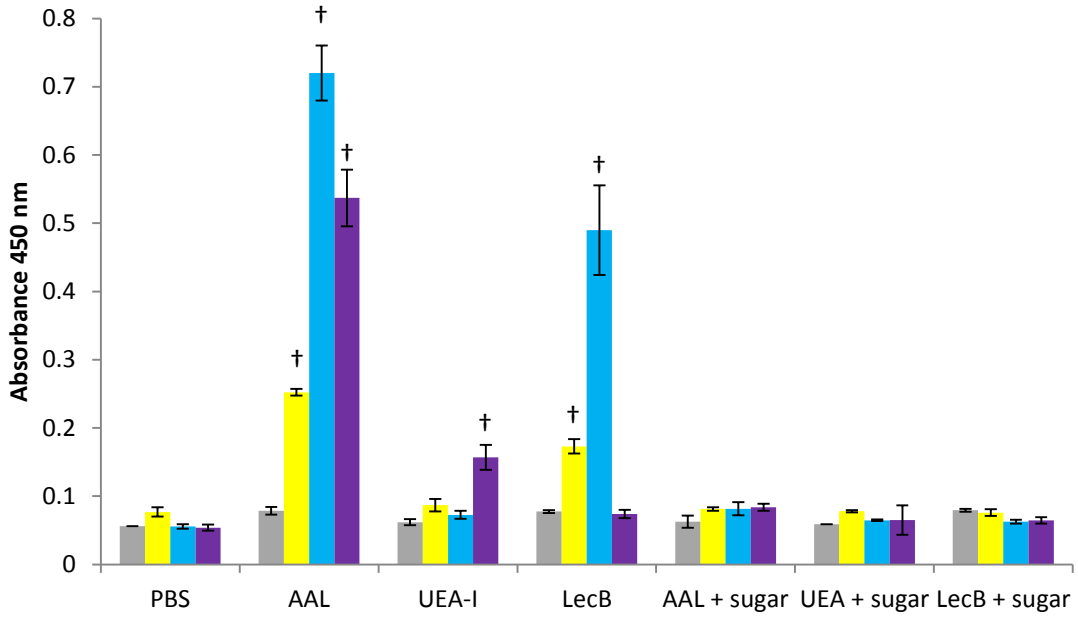


Figure 3.24: ELLA analysis of LecB compared to other fucose binding lectins AAL and UEA-I with competitive free sugar inhibition. All lectins and glycoproteins were used at 5 µg/mL. Bars on the chart are as follows; Grey = PBS, Yellow = Asialofetuin, Light blue = Thyroglobulin, Purple = Bovine Mucin. The competitive inhibiting sugar used for the three controls was 100 mM L-Fucose. † denotes a positive value surpassing the threshold limit (Section 2.30).

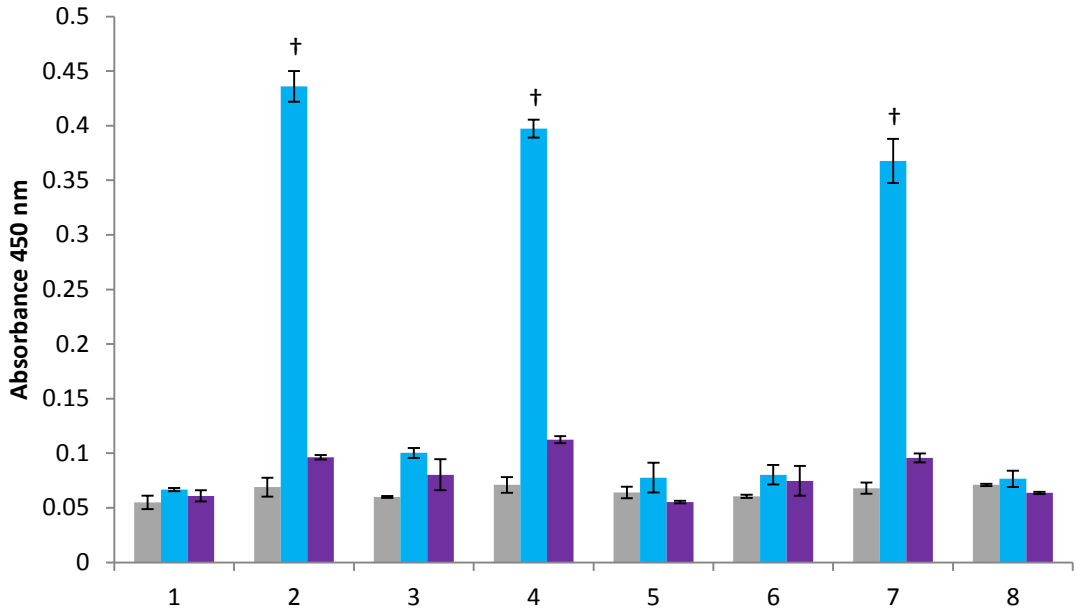


Figure 3.25: ELLA analysis of LecB binding when different competitive and non-competitive free sugars are added. Thyroglobulin was used as the positive control while bovine mucin was used as the negative control. All lectins and glycoproteins were used at 5 µg/mL. 1) PBS, 2) LecB, 3) LecB + 200 mM D-galactose, 4) LecB + 100 mM D-fucose, 5) LecB + 100 mM L-fucose, 6) LecB + 200 mM mix of α-methylmannoside and α-methylglucoside, 7) LecB + 200 mM D-glucose, 8) LecB + 200 mM D-mannose. Bars on the chart; Grey = PBS, Light blue = thyroglobulin, Purple = bovine mucin. † denotes a positive value surpassing the threshold limit (Section 2.30).

LecB is characterised as a fucose/mannose binding lectin, therefore free sugar L-fucose and D-mannose can both be used to selectively and competively inhibit LecB binding to its specific target. After LecB activity was proven, see figure 3.24, different sugars were used to highlight LecBs binding specificity. Both D-mannose and L-fucose inhibit LecBs ability to bind its target while D-glucose and D-fucose have no inhibitory effect on LecB. While LecB is characterised as a lectin that does not bind to D-galactose, D-galactose has been shown to inhibit LecB binding when used as a free sugar (Sabin et al. 2006). In Figure 3.25, 200 mM D-galactose is shown to inhibit LecB binding. Free sugar galactose has shown an ability to repeatedly inhibit LecB binding but LecB has no binding affinity for galactose in complex. In figures 3.26 and 3.27 the ability of LecB and some of its mutants to bind to mannose-linked BSA and thyroglobulin without showing any affinity for galactose-linked BSA. The original aim for LecB 3A was to knock out its binding ability by substituting some of its important binding pocket amino acid residues with alanine. These mutations have not proven to remove or reduce the proteins ability to bind to its target. Also the original aim for LecB S24T was to reduce LecBs ability to bind mannose while keeping its ability to bind to bind to fucose. These mutations again have not proven to remove or reduce the proteins affinity for mannose residues. Both of these proteins were used as probed against mannose-linked BSA as well as thyroglobulin because these are two positive controls for wild type LecB. Galactose-linked BSA was also used as a negative control. Competitive inhibition of these two LecB mutants as well as wild type LecB is also carried out to highly protein binding specificity. The comparison of LecB, LecB 3A and LecB S24T are shown in figures 3.26 and 3.27.

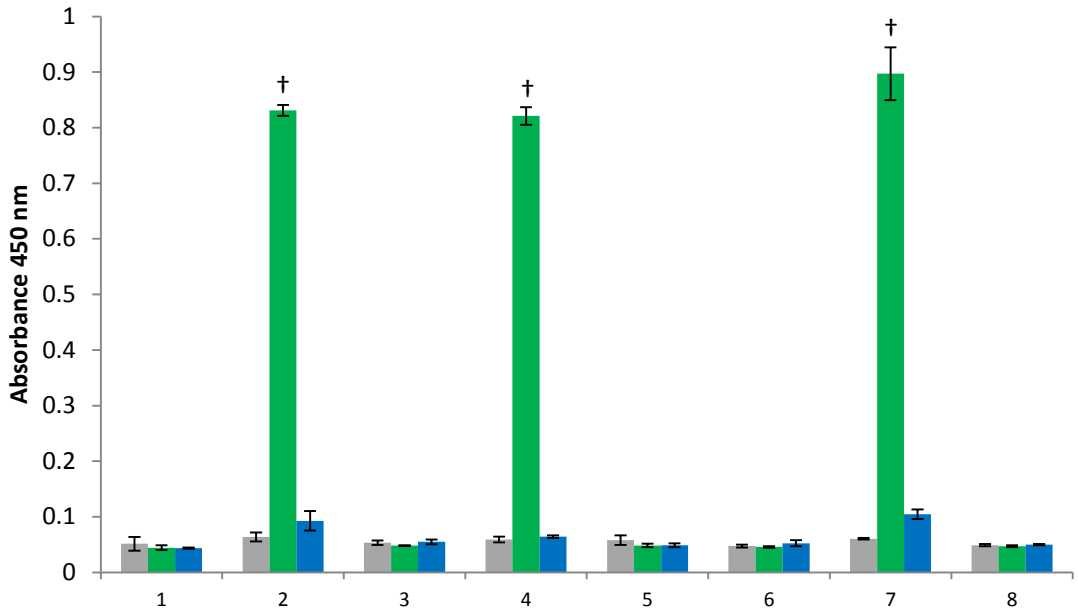


Figure 3.26: ELLA analysis of LecB mutants compared to wild type LecB using BSA-linked sugars. All lectins and glycoproteins were used at 5 µg/mL. Mannose-linked BSA is used as the positive control, whereas galactose-linked BSA is used as the negative control. LecB 3A, LecB S24T and wild type LecB were all used as probes, with 200 mM D-mannose or 100 mM L-fucose used as the competitive inhibitory sugar. 1) PBS, 2) LecB 3A, 3) LecB 3A + 200 mM D-mannose, 4) LecB S24T, 5) LecB S24T + 200 mM D-mannose, 6) LecB S24T + 100 mM L-fucose, 7) Wild type LecB, 8) Wild type LecB + 100 mM L-fucose. Bars on the chart; Grey = PBS, Green = Mannose-linked BSA, Blue = Galactose-linked BSA. † denotes a positive value surpassing the threshold limit (Section 2.30).

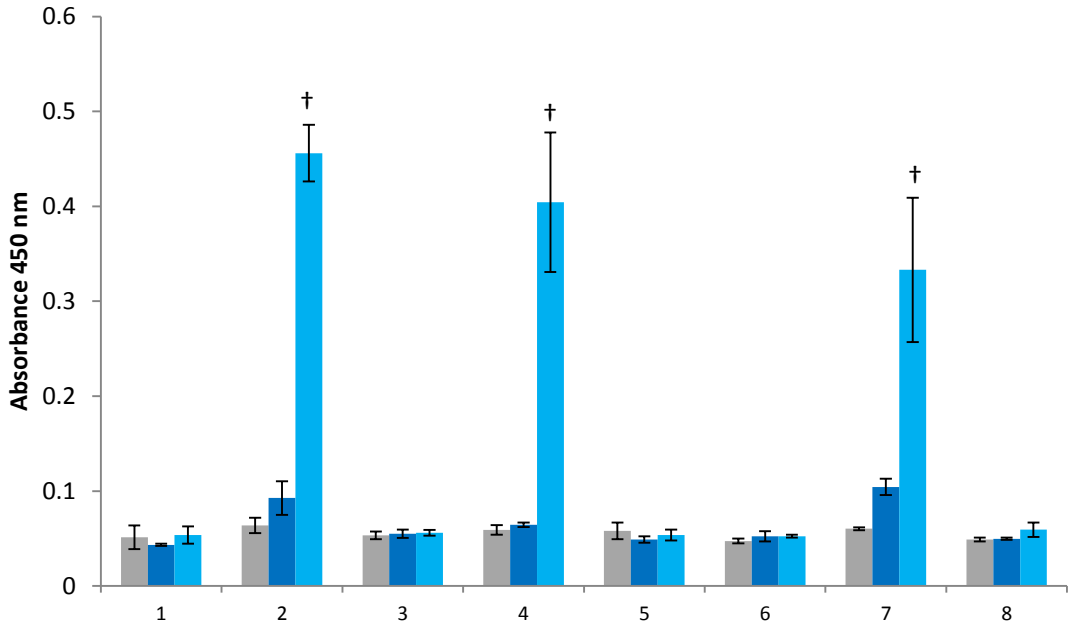


Figure 3.27: ELLA analysis of LecB mutants compared to wild type LecB using commercial glycoprotein and BSA-linked sugars. All lectins and glycoproteins were used at 5 µg/mL. Thyroglobulin is used as the positive control, whereas galactose-linked BSA is used as the negative control. LecB 3A, LecB S24T and wild type LecB were all used as probes, with 200 mM D-mannose or 100 mM L-fucose used as the competitive inhibitory sugar. 1) PBS, 2) LecB 3A, 3) LecB 3A + 200 mM D-mannose, 4) LecB S24T, 5) LecB S24T + 200 mM D-mannose, 6) LecB S24T + 100 mM L-fucose, 7) Wild type LecB, 8) Wild type LecB + 100 mM L-fucose. Bars on the chart; Grey = PBS, Dark blue = Galactose-linked BSA, Light blue = Thyroglobulin. † denotes a positive value surpassing the threshold limit (Section 2.30).

Another one of the LecB mutants generated from the previous section is LecB with three added lysine residues between the poly-histidine tag and the start codon of the LecB gene. This LecB mutant is called LecB 3K representing the three additional lysine residues. From the animo acid sequence of LecB there is one solitary lysine residue in the sequence and along with the *N*- terminus this leaves two possible sites for biotinylation. Biotin molecules are commonly added to proteins via the NH^{3+} side group of the *N*- terminus amino acid and the variable group on the amino acid lysine. To create a LecB probe that can be successfully biotinylated in the same fashion as other commercially available eukaryotic lectins and recombinant prokaryotic lectins a number of lysine residues were added to increase the amount of available biotinylation sites. Three additional lysine residues were added with the biotinylation results dictating if more lysine residues were necessary, i.e. if protein could not be successfully biotinylated more lysine residues would need to be added. The lysine residues added at that position may also impact the poly-histidine tags ability to bind to the IMAC resin. The idea to add lysine was that potential biotinylation of the protein could be increased without reducing the proteins binding affinity or ability to be purified via its poly-histidine tag. In Section 2.4, the purified version of LecB 3K was observed and the protein was biotinylated (see section 2.22). In figure 3.28 the newly produced LecB 3K and biotinylated LecB 3K were tested for binding activity compared to wild type LecB using mannose and fucose linked BSA. The binding affinity of LecB 3K was not impacted by the addition lysine residues. The biotinylated version of the protein is used as a probe with an anti-biotin antibody, instead of the usual anti-histidine antibody used for the purified recombinant lectins. The binding specificity was confirmed using competitive free sugar inhibition (see Figure 3.28).

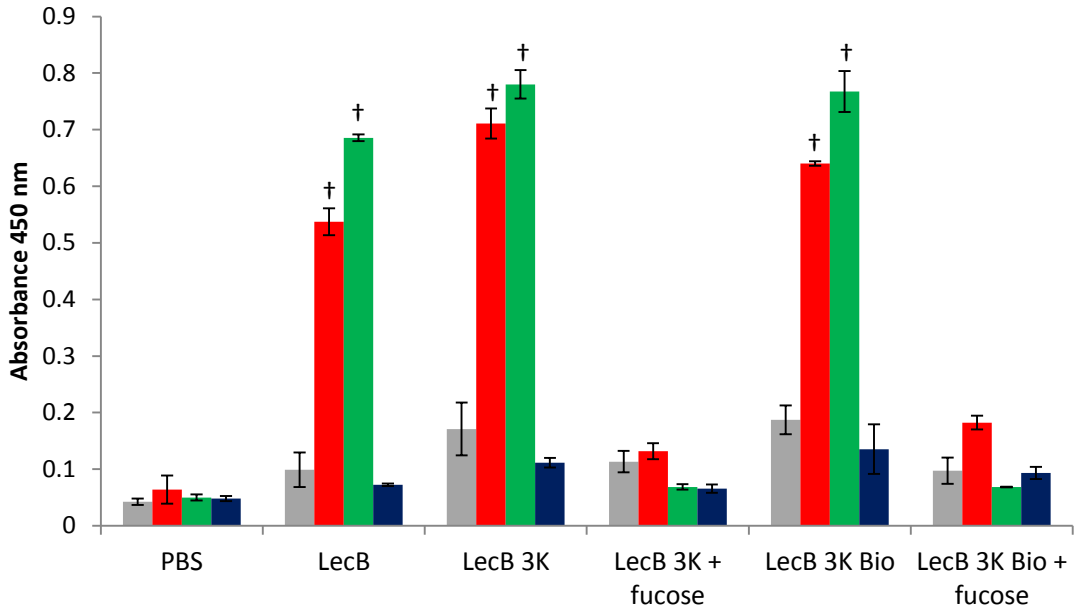


Figure 3.28: ELLA analysis of LecB 3A and biotinylated LecB 3A compared to wild type LecB with competitive inhibitory sugars. All lectins and glycoproteins were used at 5 µg/mL. The inhibitory sugar used for LecB, LecB 3K and biotinylated LecB 3K (LecB 3K bio) is 100 mM L-fucose. Bars on the chart; Grey = PBS, Red = Fucose-linked BSA, Green = Mannose-linked BSA, Navy = GlcNAc-linked BSA. † denotes a positive value surpassing the threshold limit (Section 2.30).

The last LecB variant generated in Sections 3.3 and 3.4 is the fusion protein eGFP-LecB. The enhanced version of GFP has been used as a fusion partner for a range of different applications in research. The hypothesis for creating a fusion partner protein was to create a stable eGFP-LecB fusion protein that could be used for fluorescent microscopy and flow cytometry without needing to add an additional fluorophore. The LecB gene was added to the *C*- terminus of the eGFP gene. Similar to the other LecB variants, the binding specificity and affinity were tested using ELLA analysis with commercial glycoproteins. Thyroglobulin again was used as the positive control whereas fetuin was used as negative controls.

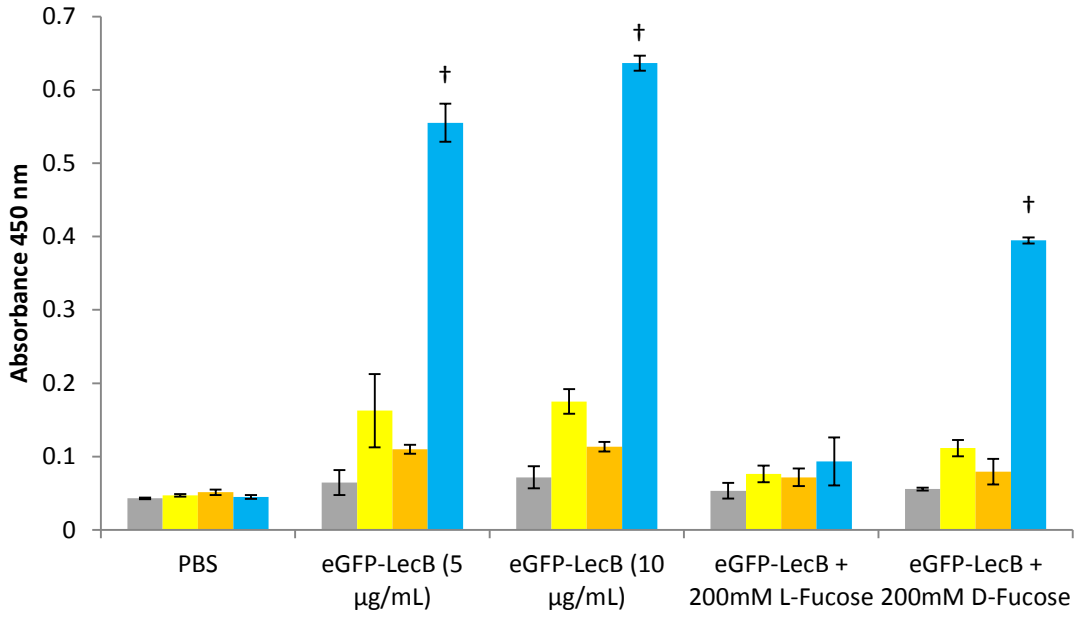


Figure 3.29: ELLA analysis of eGFP-LecB fusion protein activity using competitive and non-competitive inhibitory free sugars. Both L- and D- fucose free sugars were used to show competitive inhibition of eGFP-LecB. eGFP-LecB concentrations of 5 and 10 µg/mL were used to show binding activity. Bars on the chart; Grey = PBS, Yellow = Asialofetuin, Orange = Fetuin, Light blue = Thyroglobulin. † denotes a positive value surpassing the threshold limit (Section 2.30).

Considering eGFP-LecB is a larger protein than the other LecB mutants characterised in this section, two concentrations were tested to see if the saturation point of binding is reached at 5 µg/mL or is it closer to 10 µg/mL. There was no significant difference in binding when using a higher concentration of protein so 5 µg/mL of eGFP-LecB was used for all subsequent ELLA analysis using this protein.

3.7 Discussion

The original LecB stock contained the correct plasmid, confirmed by DNA sequence analysis which allowed this to be used as the template plasmid DNA for the remainder of the project. All subsequent PCRs were carried out using this template DNA. In this project all subsequent PCRs were whole plasmid PCRs which involves introducing a mutation to the gene and replicating the whole plasmid. The PCR product is then a linearised plasmid containing the desired mutation and this can be re-circularised using DNA ligase to create an intact plasmid for transformation. For the fusion protein the unique properties incorporated into the plasmid allowed for the gene to be excised using the restriction enzymes BamHI and HindIII. These two restriction sites are present in the multiple cloning site located at the *C*-terminus of eGFP as well as being either end of the PA-IIL gene encoding LecB. Using these two restriction enzymes the eGFP plasmid can be opened and allow the ligation of the LecB gene into the plasmid. Incorporating the LecB gene into the pQE30-eGFP plasmid removed one of the three AccI restriction sites. Restriction analysis using AccI confirmed the LecB gene had been successfully ligated into the pQE30-eGFP plasmid.

All sequences were confirmed using DNA sequencing. Each individual LecB mutant was compared to the original LecB nucleotide sequence. All DNA sequences were then run through the expasy online translate tool which allowed for the amino acid sequences to be aligned together and ensure the correct mutation had taken place at the correct position.

All of the LecB mutants created, as well as wild type LecB, were successfully purified and a very high level of purity was attained for each protein. In appendix A.1 there are some figures showing the difficulties faced in purifying certain LecB mutants. Increasing the number of wash steps with differing levels of imidazole or increasing the volume of each wash step allowed for successful purification. There was a large amount of LecB expressed when induced with IPTG and the LecB mutants were also easily overexpressed to allow for a large amount of protein to be purified. The optimisation of the purification method came from using lesser amounts of total cleared lysate and increasing the number of wash steps. The IMAC resin has a saturation point for how much poly-histidine tagged protein may be bound and this should not be exceeded or it will show an increased amount of the target protein in the flow through as well as retain a number of undesired proteins through proteins 'sticking' to each other via various bonds.

During purification of LecB and the library of LecB mutants there were significant issues with what believed to be protein aggregation. Even at relatively low protein concentrations LecB was appearing to fall out of solution when in protein friendly buffers, i.e. PBS with added NaCl or added glycerol. No protein storage solution could significantly decrease this amount of 'aggregation'. This phenomenon turned out not to be aggregation at all. LecB, as well as many other eukaryotic and prokaryotic lectins, requires a metal ion for its activity. Its metal binding pocket is so strong that while LecB is in complex with the nickel, via its poly-histidine tag, some of the nickel was also beginning to fit into its metal ion binding pocket. When the elution buffer is then added to elute off bound LecB this metal ion binding pocket interaction is not reversed and the resulting protein elution drags a small amount of resin through the filter in the IMAC column. This resin is now significantly heavier and settles at the bottom of elution tubes still in complex with the protein. The higher the concentration of protein the greater the amount of total resin in the sample which was misleading when determining the more protein in the sample the more 'aggregation' observed. This phenomenon was overcome by adding 1 mM EDTA to the elution fractions post elution and pre-buffer exchange.

EDTA can be damaging in high concentrations to proteins or if proteins are exposed to it for long prolonged periods of time. EDTA was often added to proteins for 30 minutes before buffer exchanging (see section 2.19). The buffer exchange then allowed for proteins to be separated from the nickel-resin complex as well as remove the EDTA and high imidazole from the solution.

Once each protein was successfully stored in the right buffer conditions at an acceptable concentration the samples were tested using the ELLA platform. This allowed for validation of all commercially purchased lectins as well as all of the LecB mutants produced. Some eukaryotic lectins have a very strong binding affinity for glycans in complex and cannot be fully inhibited by using various concentrations of free sugars. Jacalin and RCA which both bind differently linked galactose residues are not inhibited fully when pre-incubated with their competitive free sugar. Through using free sugars to selectively inhibit lectin binding the specificity of each lectin is characterised and the binding is confirmed to be strictly through the binding pocket-glycan interaction.

The wild type LecB and LecB mutants were characterised via their ability to bind specific glycoproteins or BSA-linked glycans. Each protein was individually characterised and validated based on its ability to bind its target glycan. The specificity is then confirmed using either 100-200 mM L-fucose or 200 mM D-mannose as a competitive inhibitory free sugar. LecB S24T had been hypothesised to have lower affinity for BSA-linked mannose or any mannose structure proven to have no fucosylation. After a number of different batches were purified and tested, the LecB S24T mutant was confirmed to bind BSA-linked mannose with the same affinity and specificity of wild type LecB. LecB S23A;S24A;G25A (LecB 3A) had been hypothesised to have a lower affinity or no affinity for either L-fucose or D-mannose due to three of the key binding residues being changed to non-polar amino acids. These three were selected base on their position in the protein. Other regions of the binding

pocket were also believed to be close to the proteins metal ion binding pocket. It was hoped that binding could be knocked out using mutations that did not affect the proteins ability to bind its desired metal ion, i.e. calcium. Again after a number of different purified fractions were tested, this LecB 3A was confirmed to have similar binding to that of wild type LecB. Both LecB S24T and LecB 3A were not used for further *in vitro* analysis based on their binding specificities.

LecB 3K and eGFP-LecB were both designed to retain their ability to bind the same specific targets as wild type LecB. Both of these proteins were also validated using the ELLA platform to measure binding specificity. In section 3.5 both of these proteins were characterised and validated compared to wild type LecB. These proteins both bound specifically to their target and were selectively inhibited using their corresponding competitive free sugar. LecB 3K was biotinylated and the biotinylated version bound to its specific targets also. The biotinylation of this protein allowed for the use of streptavidin linked dyes for fluorescent visualisation. These proteins were used in the next two chapters, 4 and 5, for *in vitro* lectin staining.

Chapter 4

Glyco-analysis of Colorectal Carcinoma Cell Lines SW480 and SW620 using Fluorescent Microscopy

4.1 Overview

This chapter describes the recombinant and native lectin probing of colorectal carcinoma cell lines SW480 and SW620. These two cell lines are stage 2 and stage 3 colorectal adenocarcinoma cell lines, respectively. Commercially purchased eukaryotic lectins and recombinantly produced prokaryotic lectins were used to create a qualitative glycan profile for each cell type that can be used to infer differences in the glycans on the surface of each cell type. Cells were probed with the lectins conjugated to fluorophores so they can be visualised by fluorescent microscopy. All lectin probes were used on non-fixed cells rather than fixed cells. Non-fixed cell analysis ensures only cell surface glycans are identified. Fixing cells causes permeabilisation that can generate false positive results in respect to surface binding (Holmes et al. 2001). Competitive free sugars were used to selectively inhibit the binding of some lectins to highlight the lectin binding specificity. The recombinant LecB 3K and eGFP-LecB that were generated in the previous chapter were used in unison with the eukaryotic lectins in glyco-profiling.

4.2 Cell surface glyco-analysis of SW480 and SW620 colorectal carcinoma cell lines.

Cells for these experiments were cultured as detailed in Section 2.25. For glyco-analysis the cells were seeded in 6-, 12- or 24-well plates with the concentration per well varying depending on the size of the well (Section 2.26). Cells were generally seeded at 50,000-100,000 cells per well with 3 mL of supplemented medium and allowed to proliferate for 48 hours. The time of incubation was decided upon based on cell confluency measured each day. Cells probed with lectin needed to be between 50-70% confluency. If the cells are lower than 50% confluency they may not be interacting with each other sufficiently and if the cells are above 70% confluency some cells may begin to show signs of early stage apoptosis which could give false results for surface lectin binding. Epithelial cells need cell to cell adhesion to ensure correct cell signaling is occurring; therefore cells incubated for 48 hours that achieve a confluency of 50-70% were believed to be the most representative for cell surface glycan patterns. Cells were trypsinised and seed and after 48 hours, at approximately 60% confluency, a cell count was carried out (see Section 2.25.3) and the cell viability always remained above 99%. The method for probing cells with lectins is described in Section 2.26. All lectins used throughout this chapter were conjugated with a dyelight-488 fluorophore which fluoresces under the FITC (green) channel on the fluorescent microscope. The eGFP-LecB fusion prokaryotic lectin was not conjugated with a fluorophore because the eGFP molecule fluoresces in the same FITC channel. For all cells to be visualised the cells are counterstained with Hoescht 33342 (NucBlue®) which fluoresces in the DAPI (blue) channel on the fluorescent microscope. The counter stain is used to show the location of the cells in the field. The nuclear stain also allows for lectin binding to be matched to cells in those locations. For each experiment a well containing cells was not probed with any lectin and this acts as a negative control. Images are taken using a bright field (White light) to show the cells location. Figure 4.1 shows healthy SW480 and SW620 cells, images are taken using white light. A nuclear counter stain is also used for these to show the nucleus in the DAPI channel. Images are also taken using the FITC channel to allow for

experimental set up to remove the potential of auto-fluorescence. The exposure time for FITC is dictated by the cells positively stained with lectin and the negative control cells. Over exposure of cells without any lectins bound to the surface can cause the cells auto-fluoresce.

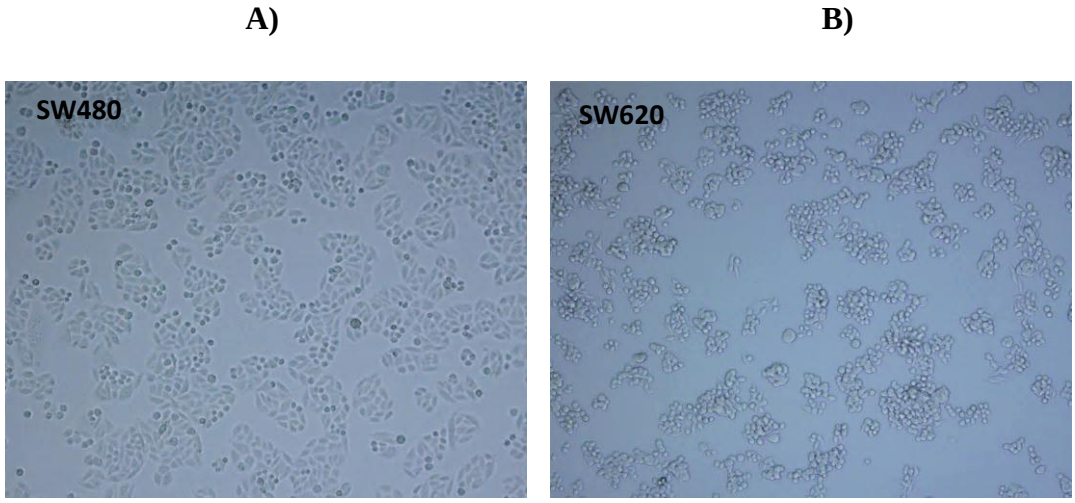


Figure 4.1: Bright field microscope image of SW480 and SW620 cells. A) SW480 at 100x magnification, B) SW620 at 100x magnification.

Cells were first probed with commercially purchased eukaryotic lectins conjugated to a fluorophore. These lectins, whose specificities were validated using ELLA analysis (see Chapter 3), have very defined activities and are required when attempting to distinguish between the two different cell types. A number of these eukaryotic lectins were used to establish a cell surface glycan profile for each of the two cell types. Each lectin used has a different specificity (see Section 2.24 for lectin specificities) and the lectins ability to bind the specific glycans on the surface of the cells will indicate the presence, or absence, of individual glycans or specific glycan linkages.

A panel of specific lectins was used to identify the presence of certain sugar residues (see table 4.1). Chapter One outlined that most eukaryotic glycans are made up of a small number of different monosaccharides. The main sugars tested for are mannose, galactose, fucose, GalNAc, GlcNAc and sialic acid (NeuNAc). To test for the presence of these sugars the panel of lectins chosen was: Con A, AAL, UEA I, Jacalin, GSL II, MAL II, WGA, SNA, HPA as well as the recombinant prokaryotic lectins eGFP-LecB and biotinylated LecB 3K. These lectins covered the majority of important sugars. The cells were grown in fetal calf serum that contains a number of glycosylated proteins that can adhere to the cell culture plates. Lectins bind to these glycosylated proteins resulting in high background fluorescence. In table 4.1 the specificities for the lectins used in this chapter are shown.

Lectin	Sugar specificity
ConA	Core Mannose
AAL	α-1-6-Fucose
UEA I	α-1-2-Fucose
Jacalin	T-antigen / β-1,3-N-acetylgalactosamine
WGA	GlcNAc, NeuNAc
GSL II	Terminal GlcNAc
MAL II	α-2,3-NeuNAc
SNA	α-2,6-NeuNAc
HPA	GalNAc, Tn antigen, GlcNAc
*Biotinylated LecB 3K	Fucose, mannose
*eGFP-LecB	Fucose, mannose

Table 4.1: Specificity of lectins used to probe SW480 and SW620 for analysis by fluorescent microscopy. *Indicates the recombinantly produced LecB 3K and eGFP-LecB from chapter 3. Eukaryotic lectins supplied by Vector Labs and Sigma-Aldrich.

4.3 Live cell glyco-analysis of SW480 and SW620 cell lines using fluorescent microscopy

The panel of lectins (Table 4.1) was used to probe SW480 and SW620 cells to determine the presence, or absence, of specific glycans or identify specific glycan linkages. To prove the binding of each lectin is specifically due to its lectin-glycan interaction, and not non-specific protein-protein interactions, free sugar in solution at defined concentrations were used to selectively inhibit lectin binding. The only free sugar not used for competitive inhibition was sialic acid for MAL II and SNA. Sialic acid has a low pH and causes damage to the cells making any subsequent results difficult to interrupt. WGA requires and acid/salt wash (as per Vector Laboratories® guidelines) to inhibit lectin binding, therefore there was no competitive free sugar used for inhibition. In table 4.2 the sugars used for each lectin to competitively inhibit their binding is shown. Not all eukaryotic lectins can be inhibited fully using free sugars (see Chpater 3). Jacalin is an example of this and in figure 4.4 jacalin binding is still observed even in the presence of a high concentration of free sugar. The majority of these lectins were fully inhibited using their corresponding free sugar. All images are either taken at 100x magnification or 400x magnification. Some images obtained contain high level non-specific background binding either caused by lectins binding to glycosylated serum proteins or cellular debris. All images taken for lectins that exhibit high background are taken at 100x magnification for clearer images.

Lectin	Competitive inhibitory free sugar
ConA	200 mM α-methylmannoside & α-methylglucoside
AAL	100 mM L-Fucose
UEA I	100 mM L-Fucose
Jacalin	800 mM D-galactose or 100 mM Melibiose
WGA	N/A
GSL II	N/A
MAL II	N/A
SNA	N/A
HPA	200 mM *N*-acetylglucosamine or 200 mM *N*-acetylgalactosamine
Biotinylated LecB 3K	200 mM L-Fucose
eGFP-LecB	200 mM L-Fucose

Table 4.2: List of free sugars used to selectively inhibit each lectin. GSL II did not bind to either cell type so it was not necessary for an inhibitory sugar. MAL II/SNA both bind sialic acid and this free sugar is not used for inhibition studies. WGA requires acid wash step deemed too damaging for cells to carry out. The inhibitory sugars and their concentrations are supplied by Vector Laboratories® or were determined by ELLA testing in Chapter 3.

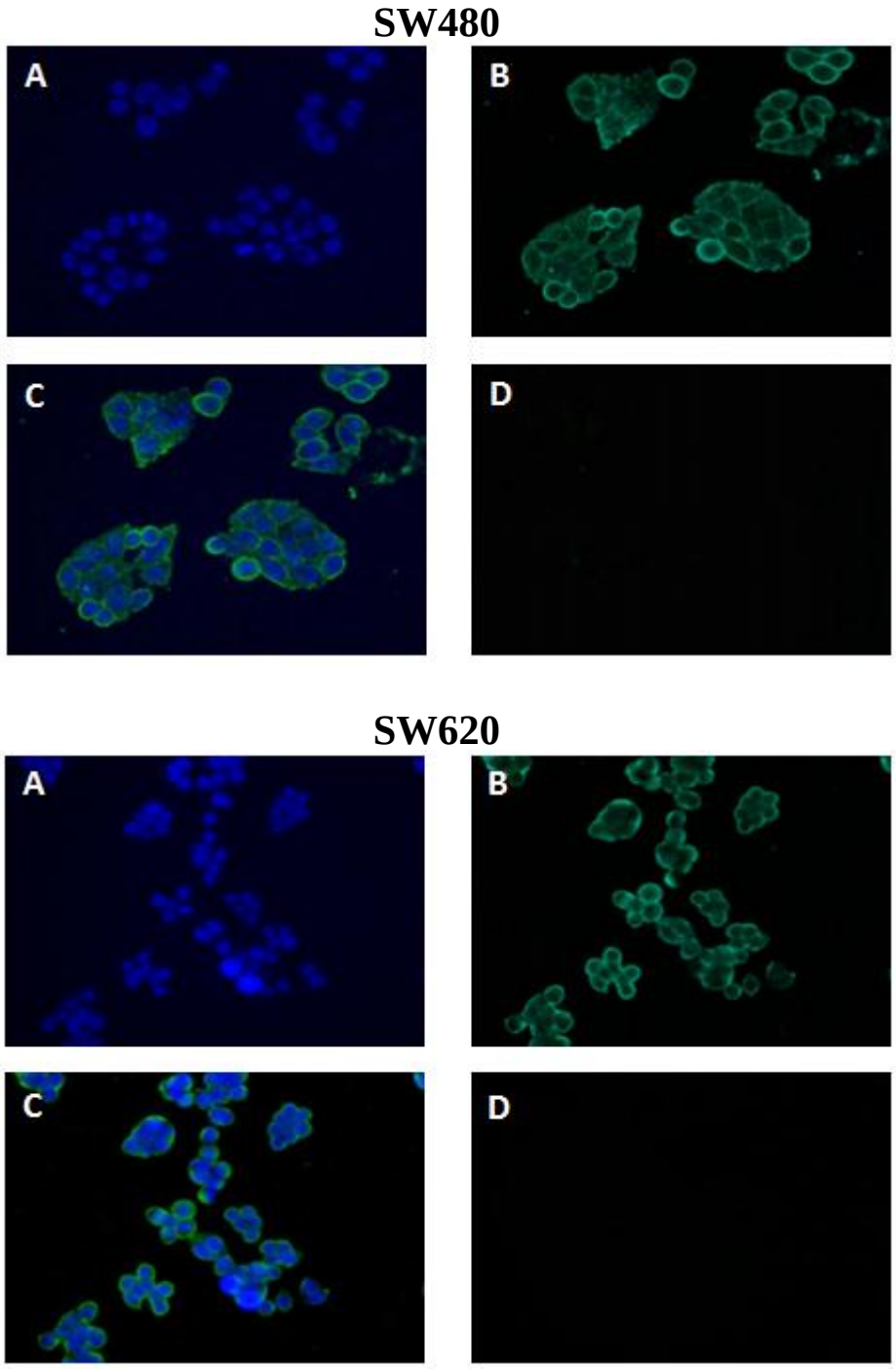


Figure 4.2: SW480 and SW620 cells probed with lectin AAL and AAL + 100 mM L-fucose. A) NucBlue nucleus stain at 400x magnification, B) AAL with dyelight-488 at 400x magnification, C) Merged image of A and B, D) AAL with dyelight-488 with 100 mM L-fucose at 400x magnification.

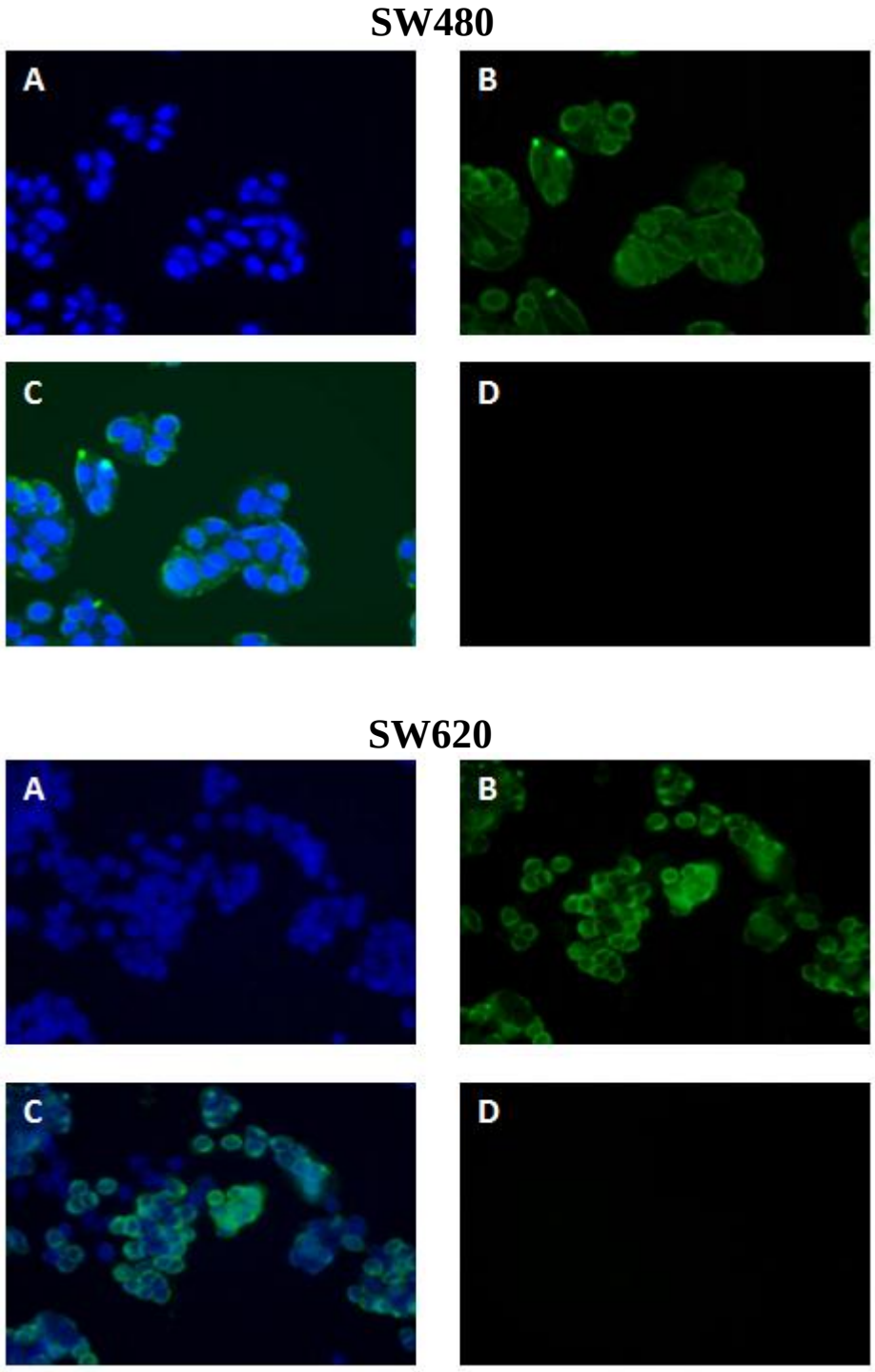


Figure 4.3: SW480 and SW620 cells probed with lectin UEA-I and UEA-I + 100 mM L-fucose. A) NucBlue nucleus stain at 400x magnification, B) UEA-I with dyelight-488 at 400x magnification, C) Merged image of A and B, D) UEA-I with dyelight-488 with 100 mM L-fucose at 400x magnification.

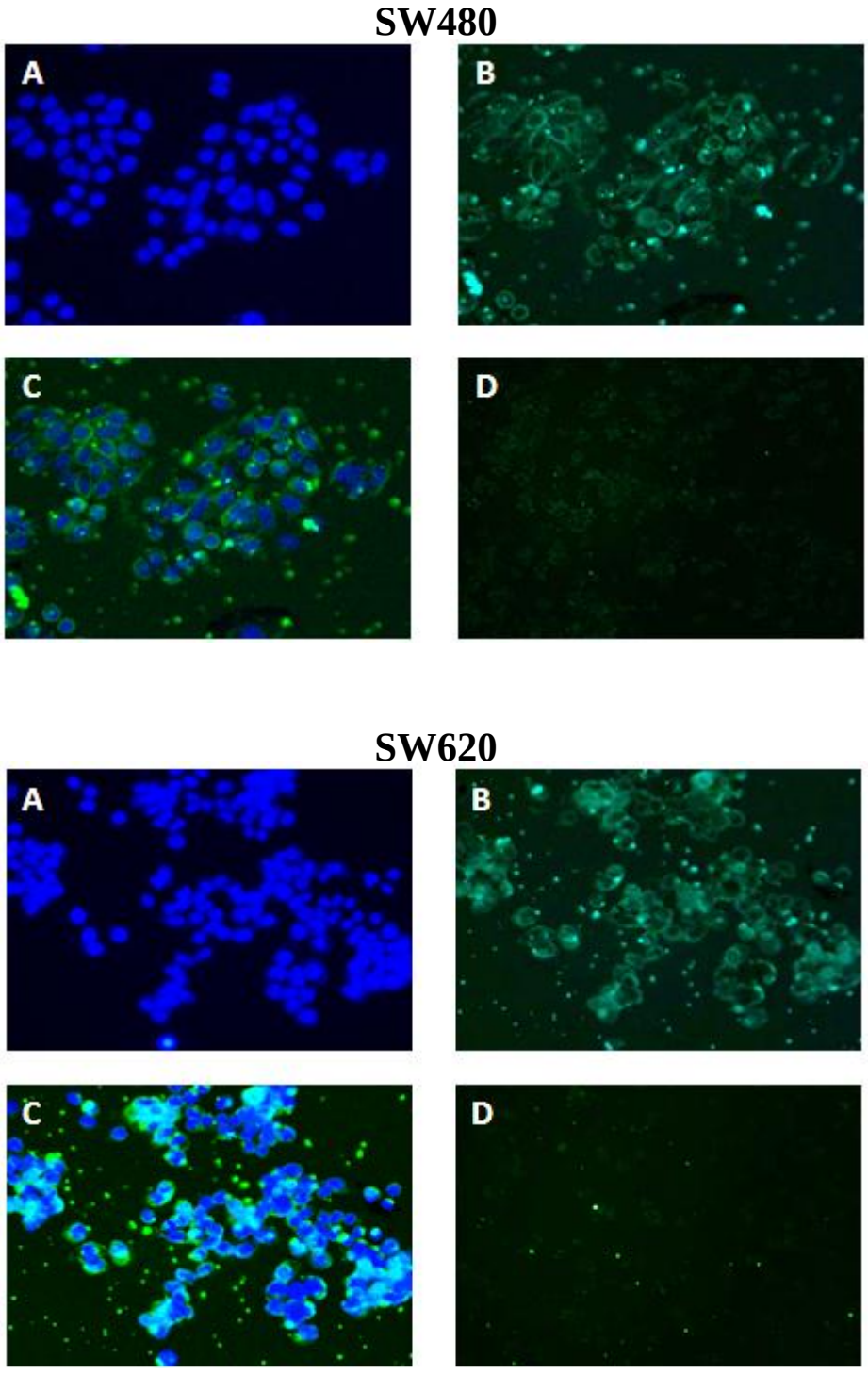


Figure 4.4: SW480 and SW620 cells probed with lectin Jacalin and Jacalin + 100 mM Melibiose. A) NucBlue nucleus stain at 400x magnification, B) Jacalin with dyelight-488 at 400x magnification, C) Merged image of A and B, D) Jacalin with dyelight-488 with 100 mM Melibiose at 100x magnification.

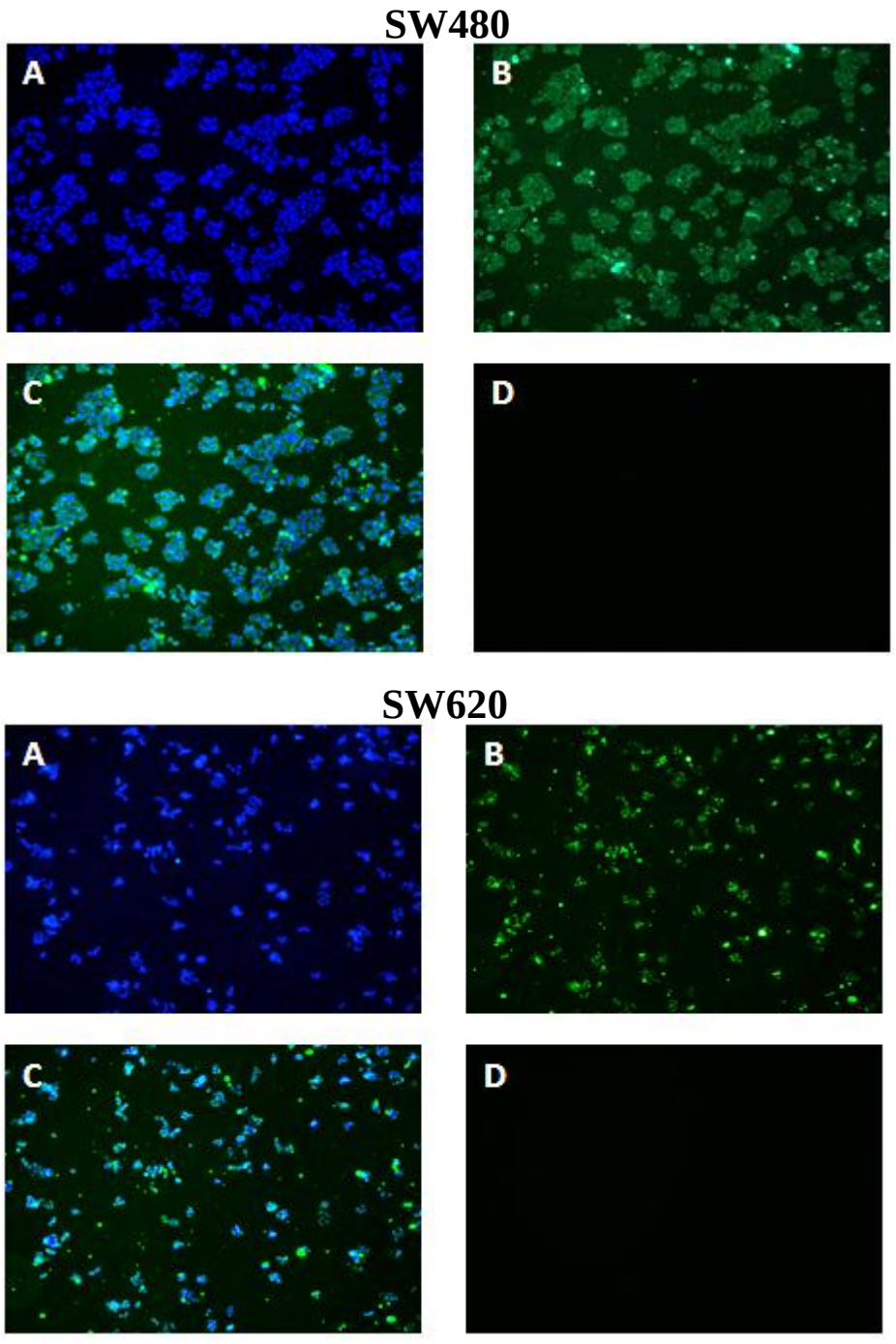


Figure 4.5: SW480 probed with lectin Con A and Con A + 200 mM α-methylmannoside / α-methylglucoside. A) NucBlue nucleus stain at 100x magnification, B) Con A with dyelight-488 at 100x magnification, C) Merged image of A and B, D) Con A with dyelight-488 with 200 mM α-methylmannoside / α-methylglucoside at 100x magnification.

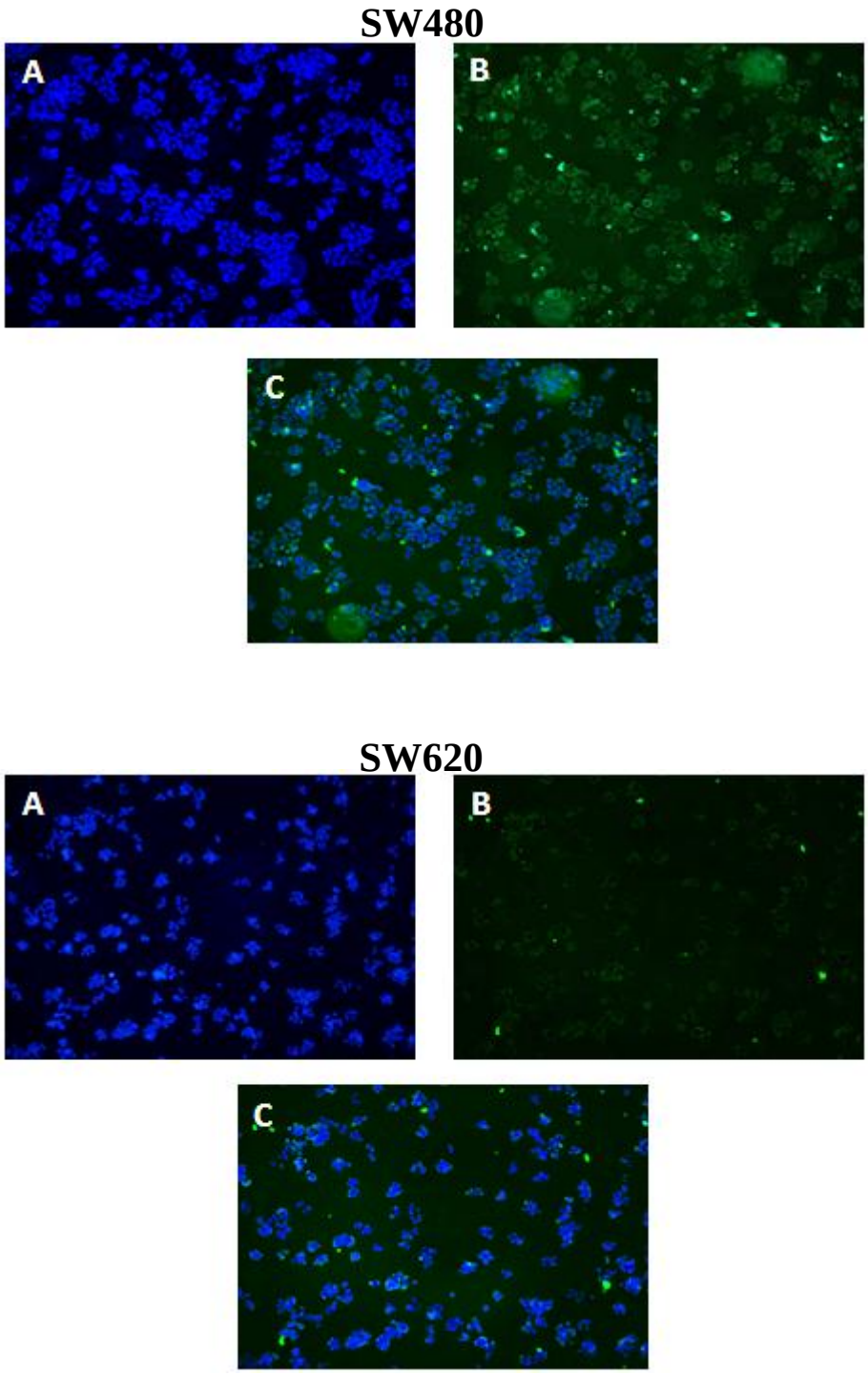


Figure 4.6: SW480 and SW620 cells probed with lectin Mal II. A) NucBlue nucleus stain at 100x magnification, B) Mal II with dyelight-488 at 100x magnification, C) Merged image of A and B.

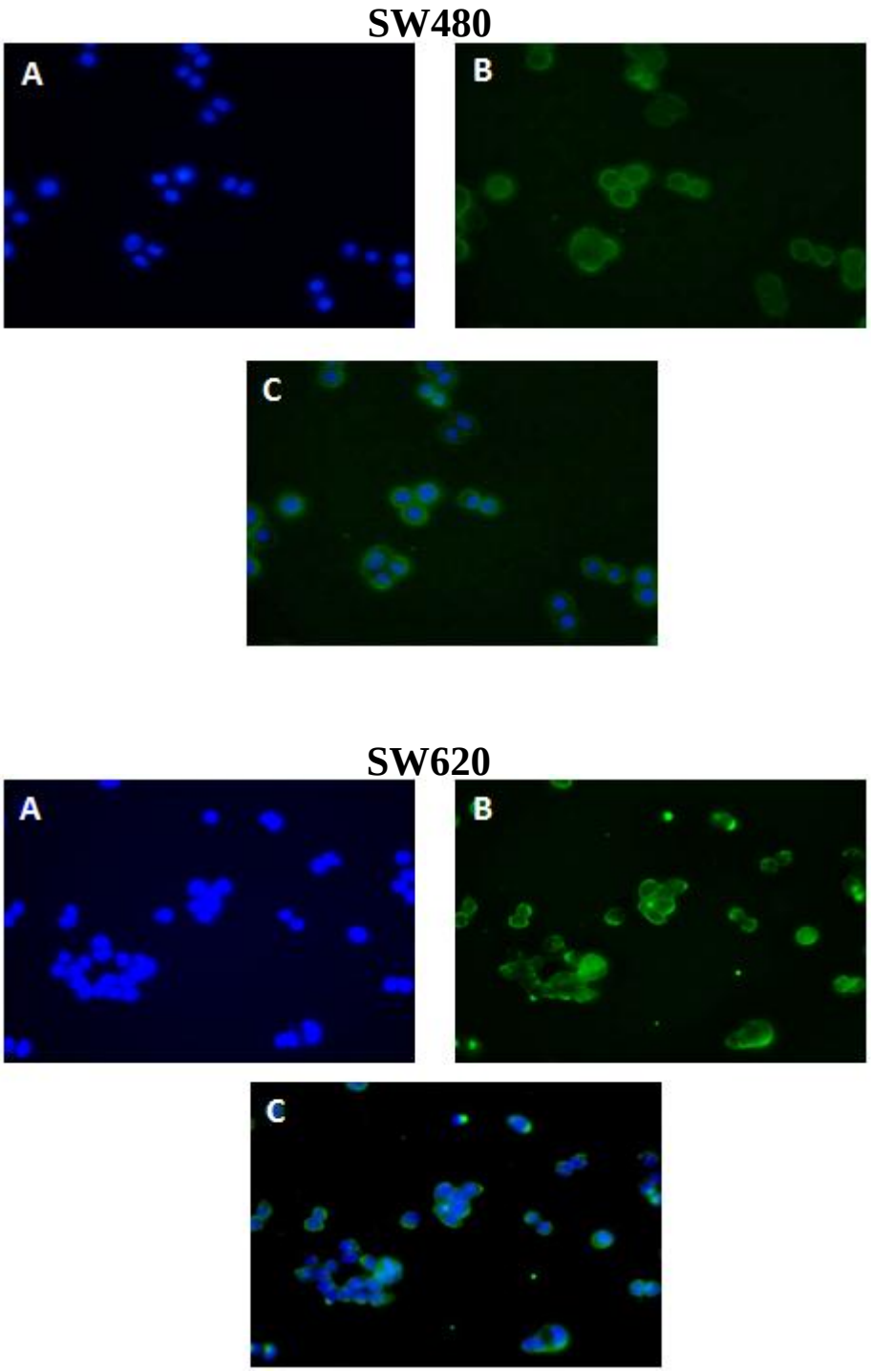


Figure 4.7: SW480 and SW620 cells probed with lectin WGA. A) NucBlue nucleus stain at 400x magnification, B) WGA with dyelight-488 at 400x magnification, C) Merged image of A and B.

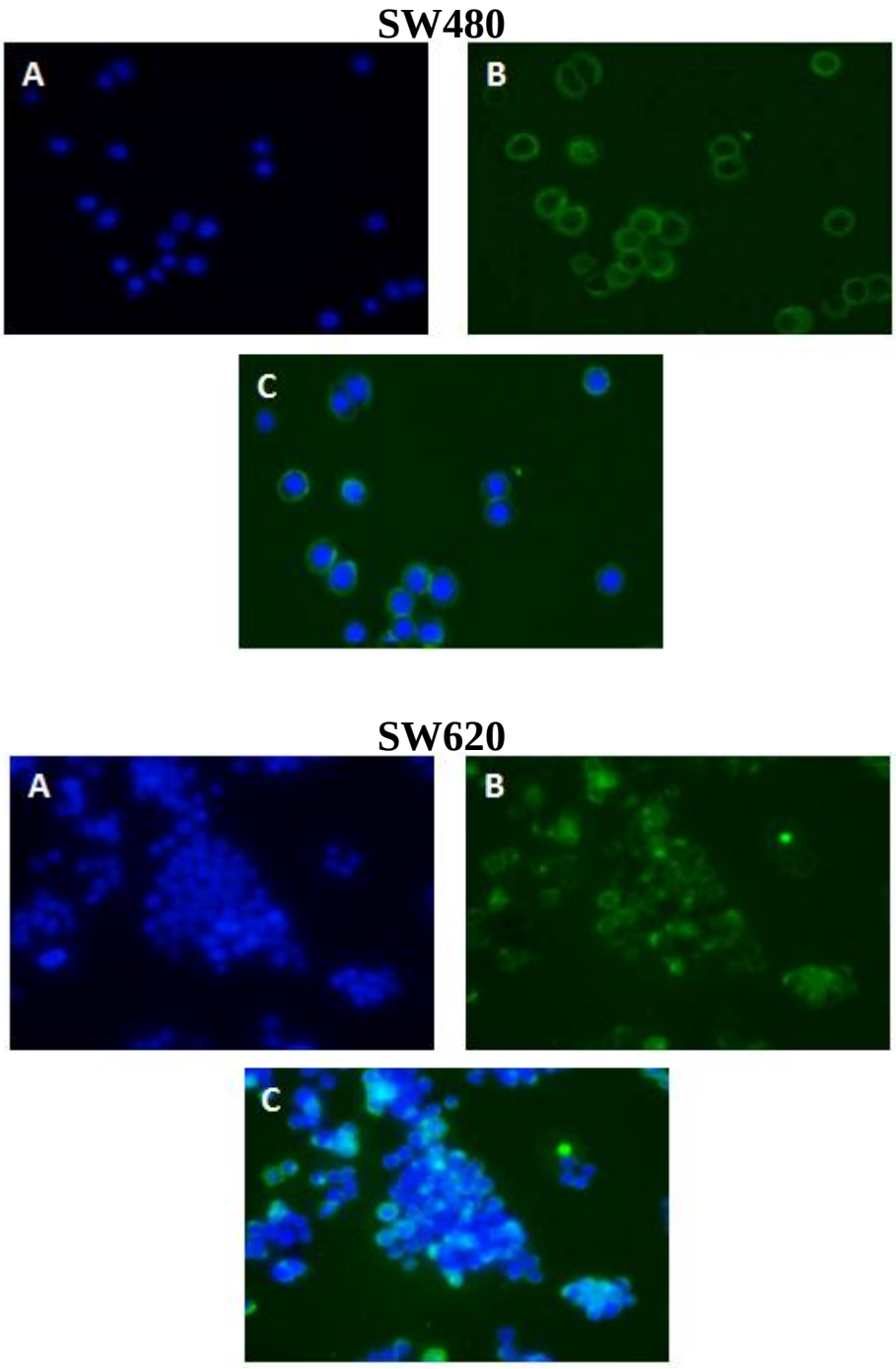


Figure 4.8: SW480 and SW620 cells probed with lectin SNA. A) NucBlue nucleus stain at 400x magnification, B) SNA with dyelight-488 at 400x magnification, C) Merged image of A and B.

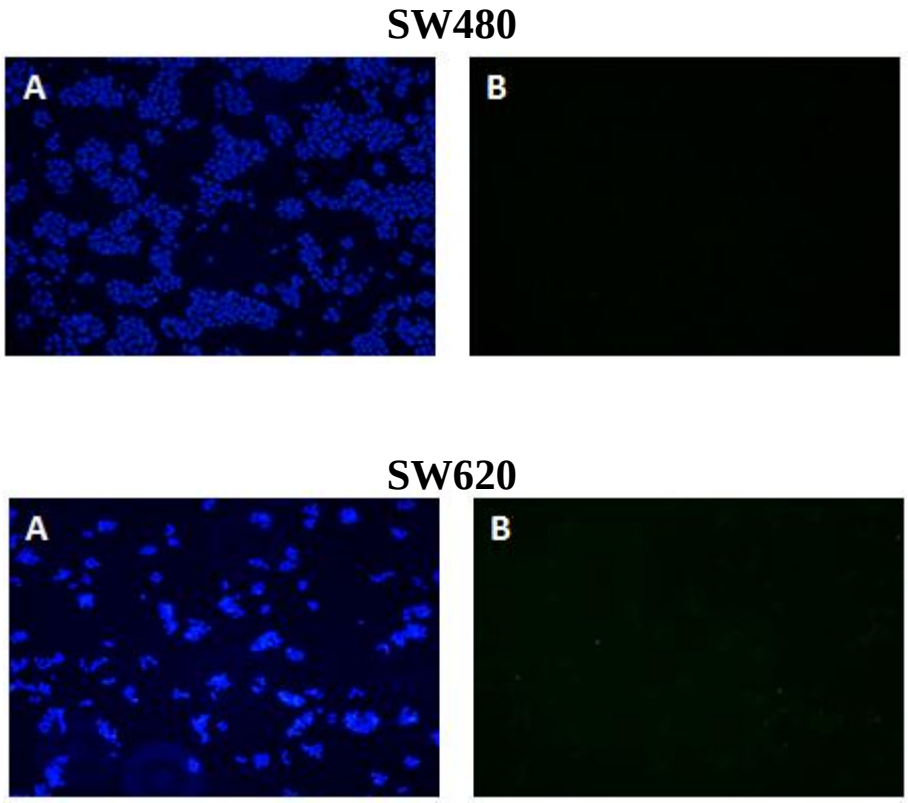


Figure 4.9: SW480 and SW620 cells probed with lectin GSL II. A) NucBlue nucleus stain at 400x magnification, B) GSL II with dyelight-488 at 400x magnification.

GSL II was used as a negative control for both cell lines as healthy cells are not known to have exposed terminal GlcNAc on their cell surface (Franz et al. 2006). The eukaryotic lectins used to probe live cells give an indication of the specific glycans and glycan linkages present on the surface of SW480 cells. The recombinant prokaryotic lectins generated in this project (see Chapter 3) were also used to identify specific glycans. In figures 4.10 and 4.11 the biotinylated LecB 3K and eGFP-LecB are shown to bind to SW480 and SW620. AAL and UEA-I binding showed that there were fucose residues of differing linkages on the surface of SW480 and the LecB variants are used to back up this claim and validate the use of these recombinantly produced lectins in live cell analysis.

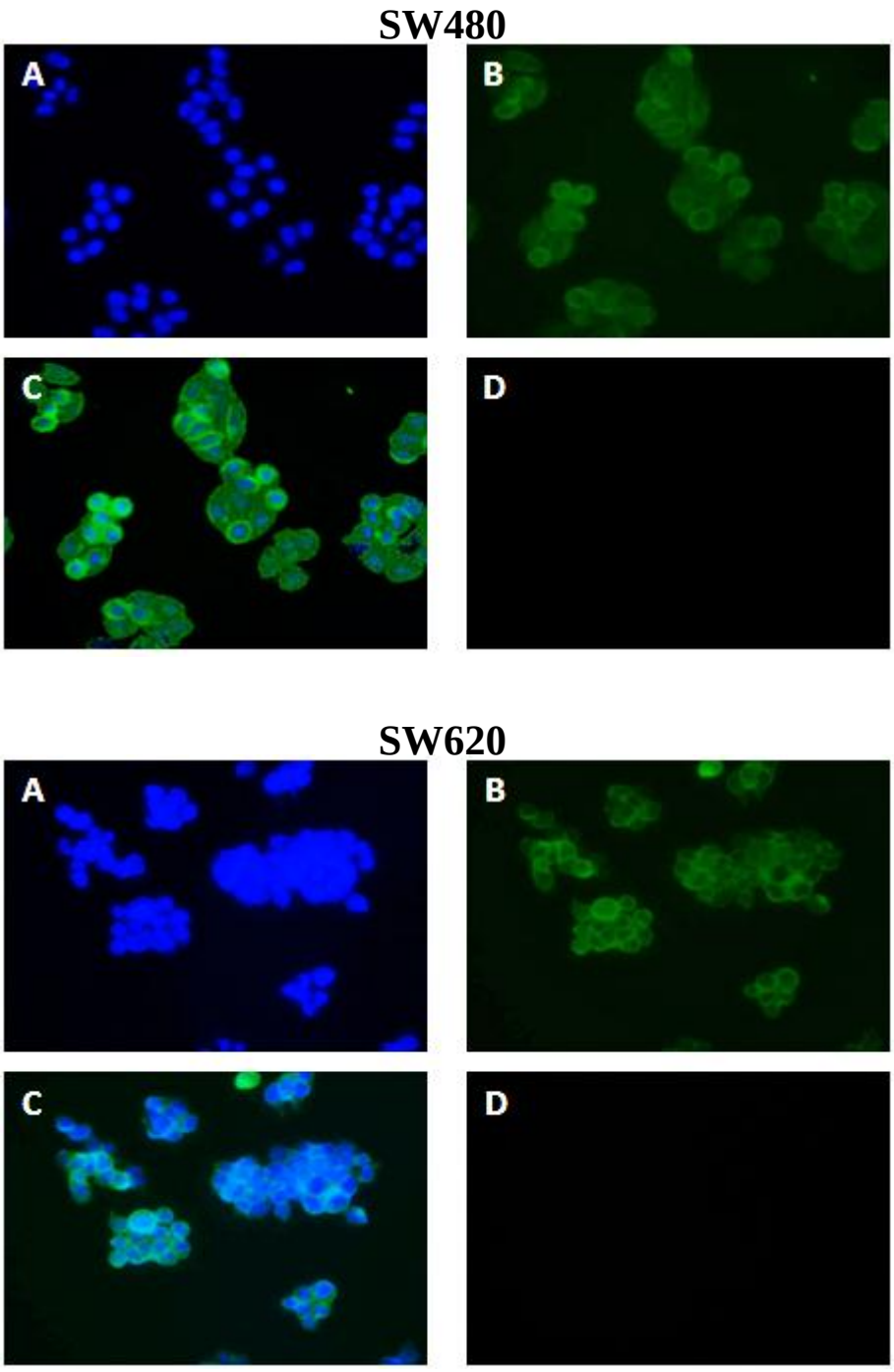


Figure 4.10: SW480 and SW620 cells probed with lectin biotinylated LecB 3K and biotinylated LecB 3K + 200 mM L-fucose. A) NucBlue nucleus stain at 400x magnification, B) LecB 3K with dyelight-488 at 400x magnification, C) Merged image of A and B, D) LecB 3K with dyelight-488 with 200 mM L-fucose at 400x magnification.

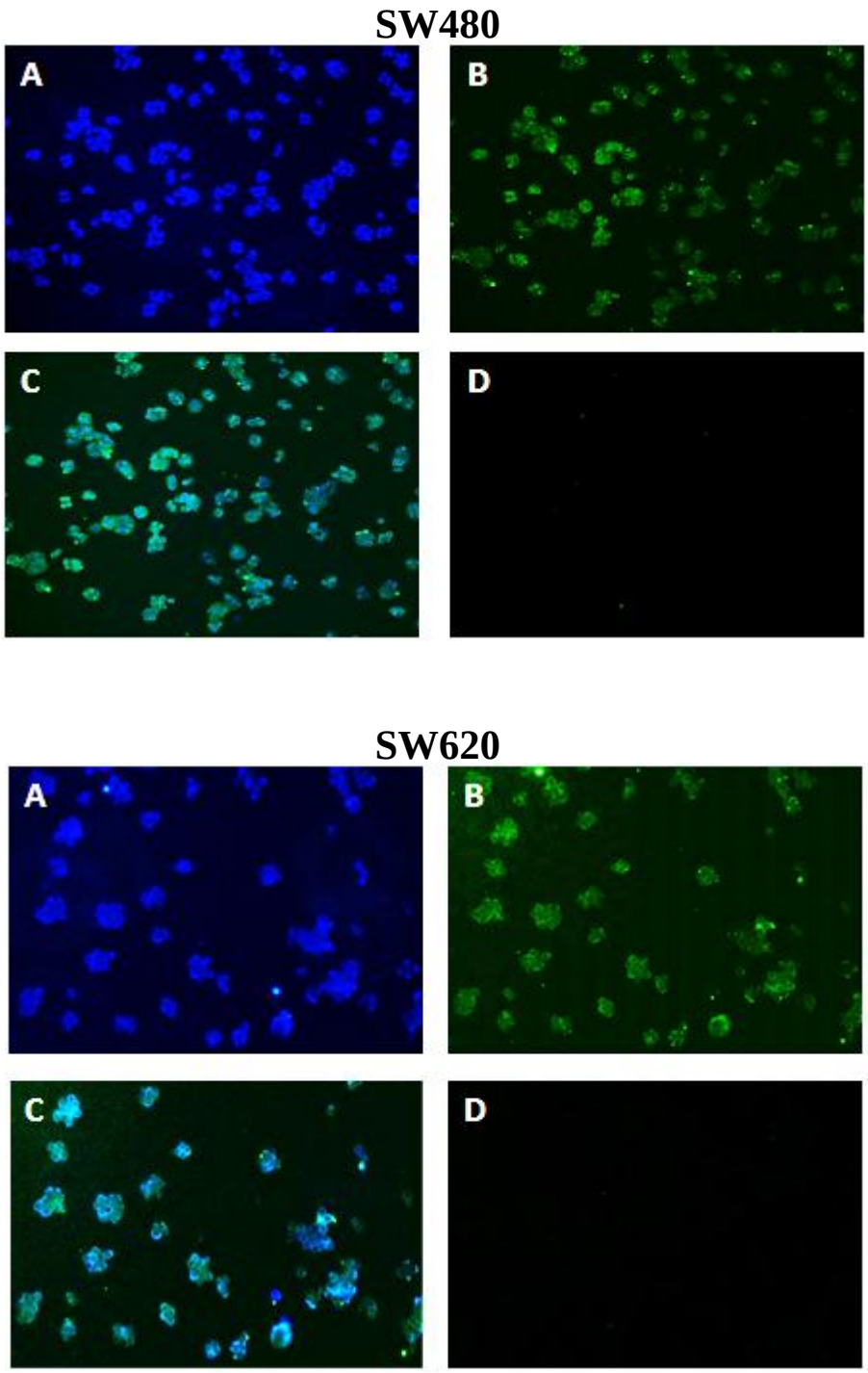


Figure 4.11: SW480 and SW620 cells probed with lectin eGFP-LecB and eGFP-LecB + 200 mM L-fucose. A) NucBlue nucleus stain at 400x magnification, B) eGFP-LecB with dyelight-488 at 100x magnification, C) Merged image of A and B, D) eGFP-LecB with dyelight-488 with 200 mM L-fucose at 100x magnification.

4.4 Analysis of Helix Pomatia Agglutinin (HPA) binding to SW480 and SW620 cells using fluorescent microscopy

One lectin used that shows interesting binding results was the *Helix Pomatia* agglutinin (HPA) lectin. Other studies highlighted HPA as a lectin that could distinguish between the non-metastatic SW480 cells and the metastatic SW620 cells. HPA was expected to bind a subset of the SW620 cells with a high level of fluorescence exhibited whereas it was expected to not bind to SW480 cells (Schumacher and Adam 1997). HPA has been used to distinguish between metastatic and non-metastatic cells using both SW480 and SW620 cells (Saint-Guirons et al. 2007; Peiris et al 2012). When the cells were probed with HPA and viewed using fluorescent microscopy the SW480 cells showed a low level of HPA binding and the SW620 cells showed some cells bound with HPA with a larger portion of the cells not showing HPA binding. In figure 4.12 the HPA binding of SW480 and SW620 is shown. One difference noted when comparing these results to the literature that shows no HPA binding to SW480 was that the cells were fixed to the plate before being viewed using fluorescent microscopy (Peiris et al. 2012). This experiment was repeated for SW480 cells that were paraformaldehyde fixed to the 6-well plate used to culture the cells. In figure 4.13 the results of this experiment are shown.

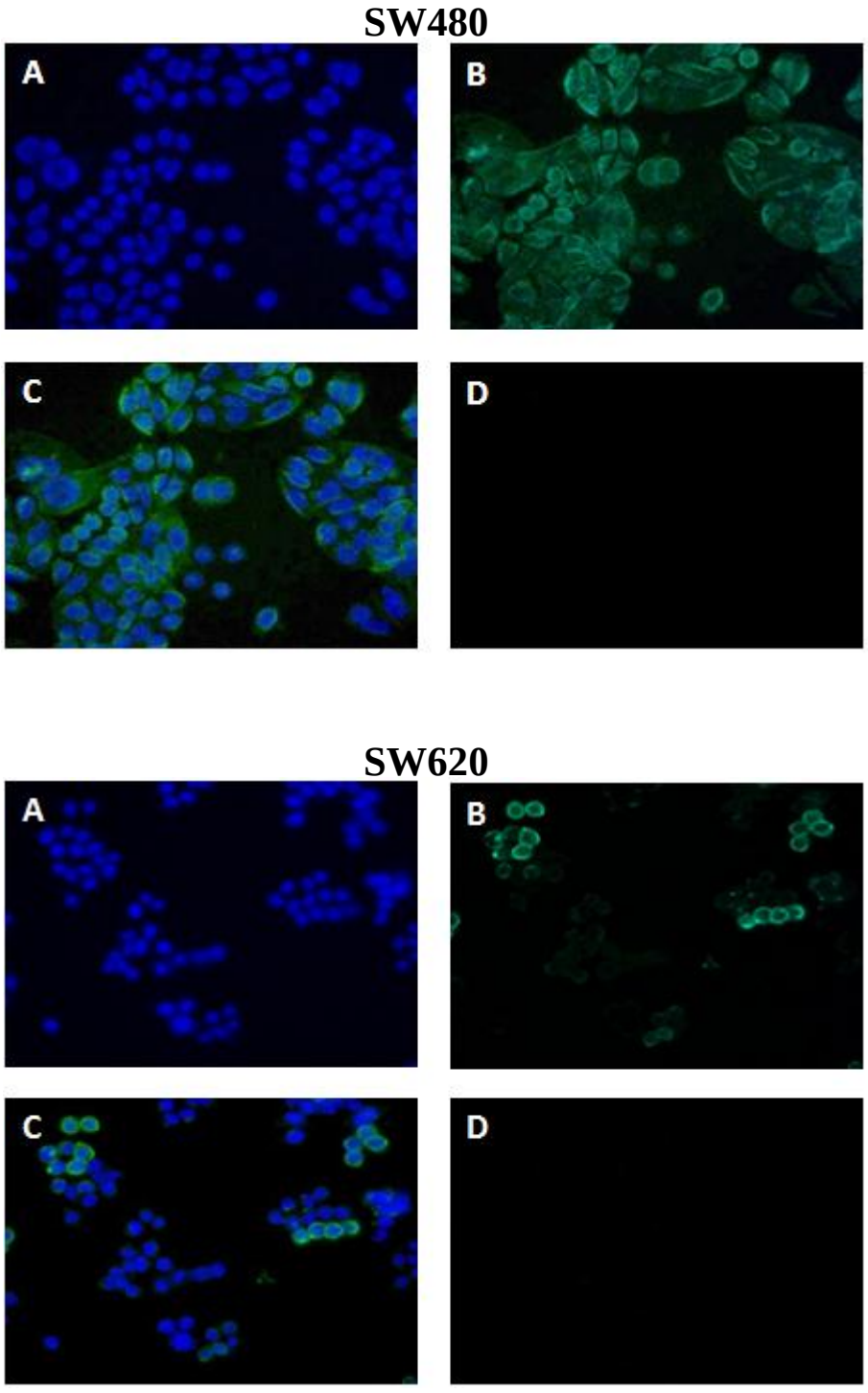


Figure 4.12: SW480 and SW620 cells probed with lectin HPA and HPA + 200 mM GlcNAc. A) NucBlue nucleus stain at 400x magnification, B) HPA with dyelight-488 at 400x magnification, C) Merged image of A and B, D) HPA with dyelight-488 with 200 mM GlcNAc at 400x magnification.

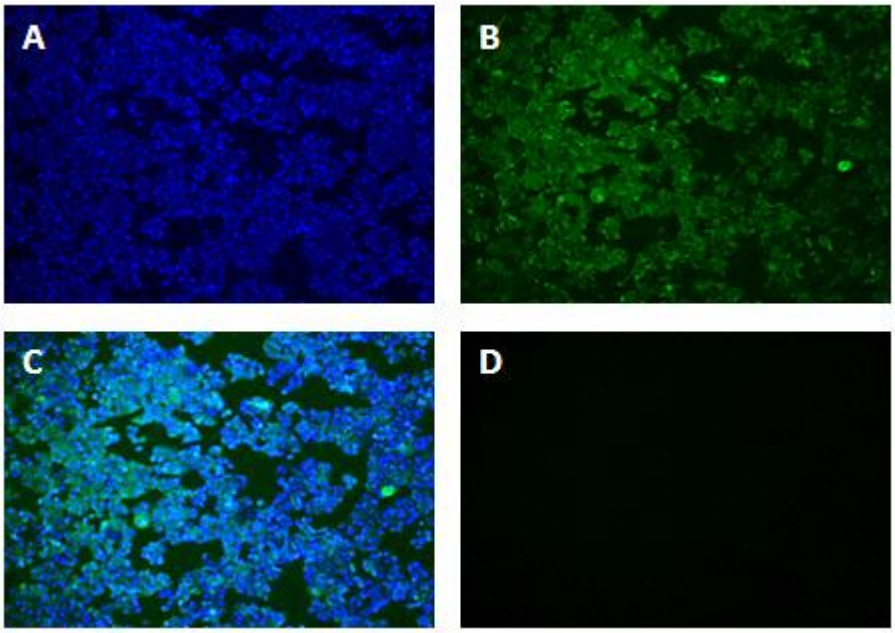


Figure 4.13: Paraformaldehyde fixed SW480 cells probed with lectin HPA and HPA + 200 mM GlcNAc. A) NucBlue nucleus stain at 100x magnification, B) HPA with dyelight-488 at 100x magnification, C) Merged image of A and B, D) HPA with dyelight-488 with 200 mM GlcNAc at 100x magnification.

4.5 Discussion

Live cell staining was carried out using eukaryotic and prokaryotic lectins to identify the glycans on the surface of SW480 and SW620 cells. Live cell analysis of both these cell types revealed the complexity of the surface glycome of mammalian cells, by showing the variety of different glycan residues present on the surface of the cells. Through using the panel of lectins chosen in Section 4.2 a select number of these sugar molecules can be detected on the surface. These lectins also identified different O- and N- glycosyl-linkages found on the surface of the SW480 and SW620 cells.

All the lectins chosen, with the exception of GSL II, were predicted to bind to these cell lines to varying degrees. GSL II binds to terminal GlcNAc which is predominantly the first sugar molecule involved in N-glycan synthesis. There is no evidence in the literature to suggest that these cancer cells would have terminal GlcNAc on the surface and this is further proven by the analysis carried out in this chapter. Cells believed to be undergoing late apoptosis can have exposed terminal GlcNAc but dead/dying cells are removed through various wash steps before lectin probing (Ise et al. 2012).

Competitive free sugar inhibitions were carried out in order to prove the selectivity of the lectin binding to the cell surface. The majority of the lectins used can be selectively inhibited by using their corresponding free sugar, whereas some like Jacalin cannot be fully inhibited using free monosaccharides. Jacalin binding is decreased in the presence of its corresponding free sugar as shown in chapters 3 and 5 as well as figure 4.4. Through lectin inhibition studies the binding specificity of each lectin was validated. As some of the lectin probes are large proteins that could weakly bind to cells through various other protein-protein interactions, the use of free sugar to competitively inhibit the lectin binding proves the lectins are only binding through their unique glycan binding site or sites. As explained in section 4.2, not all lectins could be selectively inhibited using competitive free sugars. Live cell analysis needed to be carried out at a neutral pH to ensure the cells remained viable throughout the probing process. The use of an acidic wash to inhibit WGA binding or using sialic acid, which lowers the buffer

pH, to inhibit SNA or MAL II binding causes the cells to die and potentially lead to altered binding.

As shown in figures 4.4 and 4.5 with Jacalin and Con A, serum proteins bound to the plate can cause significant background fluorescence. The plates used are treated to allow for surface proteins of adherent cells to attach which also allows glycosylated serum proteins to also attach to the surface (Zheng et al. 2006). The qualitative analysis of these lectins binding to the cell surface can still be observed but are impacted by these background lectin-serum protein interactions. This background binding is one of the significant drawbacks observed while carrying out live cell microscopy. In chapter 5 when viewing cells using flow cytometry, the cells are washed and all serum proteins are removed. In this chapter the necessity to leave cells attached to the plate for probing and incubate the cells with serum proteins that also bind to the plate causes this method to be largely qualitative and makes any quantification difficult. Live cell analysis does not allow for fixation that would allow for additional wash steps and blocking that can reduce this background fluorescence. As mentioned previously cell fixation can cause issues with lectin binding to internal glycosylated proteins (Holmes et al. 2001). Other studies have used confocal microscopy rather than routine fluorescent microscopy due to the increased power of confocal microscopy to identify lectin binding (Peiris et al. 2012; Saint-Guirons et al. 2007). Confocal microscopy allows for more quantitative analysis of lectin binding as well as highlighting lectin localisation.

There were a number of lectins used to determine the presence of glycans on the surfaces of each cell type. The main differences observed between the SW480 and SW620 cell types are the binding of MAL II and HPA. MAL II is an α-2,3-NeuNAc (sialic acid) binder. SNA is an α-2,6-NeuNAc (sialic acid) binder. Visually there appears to be similar binding of SNA to both cell types but also a decrease in the binding of MAL II to the surface of SW620 when compared to SW480. These comparisons were based off relative fluorescence for both images taken under the same conditions. No quantitative differences were observed in this chapter as the experiments were designed for qualitative results. LecB 3K and eGFP-LecB are shown in chapter 3 to bind to fucose

and mannose residues on mannose/fucose residues linked to BSA. For use in *in vitro* cellular analysis the probes needed to be tested against both cell types. Through the binding observed in figures 4.2, 4.3 and 4.5 the cells had already shown to have fucose and mannose present on the cell surface. Both LecB 3K and eGFP-LecB bind to these residues and are validated for *in vitro* cell binding by qualtitatively showing binding and competitive inhibition. The positive binding of LecB 3K to SW480 and SW620 allowed for it to be added to the panel of lectins used in Chapter 5 for flow cytometric analysis.

HPA binding observed using both fixed and non-fixed SW480 cells was unexpected. In section 4.4 the HPA lectin was identified previously as a lectin that could determine between SW480 and SW620 as it would only bind to metastatic cells, i.e. the SW620 cells. The binding profile of SW620 cells probed with HPA was as expected and in both cases the lectin is fully inhibited by its corresponding free sugar. HPA could be inhibited by both GlcNAc and GalNAc, as is shown in Section 5.5, and GlcNAc was used to competitively inhibit HPA in this chapter. The differential binding observed between both SW480 and SW620 still indicated that HPA could potentially indicate quatitative differences between the two cell types, even if the original expectations of HPA binding were not observed. A number of control experiments were designed for testing HPA binding to both cell lines and these will be explained and the results shown in chapter 5.

Due to the qualitative nature of this analysis, these two lectins, MAL II and HPA, were identified as lectins of interest for the next chapter which attempts to use flow cytometry for more quantitative analysis of lectin binding using a larger panel of lectins to test SW480 and SW620.

Establishing a glyco-profile using only ten different lectins and fluorescent microscopy is difficult. The panel chosen gives insights into the types of glycans on the surface and what sugar residues are present terminally. The negative control of GSL II gives more confidence in the binding observed with the other lectins, especially those who cannot be inhibited or full inhibited with a competitive free sugar. Chapter 5 will show more lectins that don’t bind to, or that do not bind all of, the cells when probed. The aim of

this chapter was to beging to characterise the two cell types based on the glycans they had present on their cell surface as well as indicate which lectins may show differences in the levels of binding observed.

Chapter 5

Glyco-analysis of Colorectal Carcinoma Cell Lines SW480 and SW620 using Flow Cytometry

5.1 Overview

This chapter describes the lectin probing of colorectal carcinoma cell lines SW480 and SW620. Throughout this project the two cell lines are stage 2 and 3 colorectal adenocarcinoma cell lines, respectively. Eukaryotic lectins and recombinantly produced lectins were used to probe live cells (non-fixed) via fluorescent microscopy and flow cytometry (Chapter 4). Fluorescent microscopy is a largely qualitative form of glyco-analysis, whereas flow cytometry provides more quantitative data, allowing for specific changes in lectin binding between the two cell lines to be highlighted and compared. The binding analysis of live cells provides knowledge on what type of glycans are present on the cell surfaces of each cell line. Through the use of different lectins, which bind different target glycans, the presence or absence of specific glycans and glycan linkages can be observed. Additionally, the fold increase or decrease of specific glycans can also be observed. Competitive free sugars, as well as non-competitive free sugars are used to validate the lectin binding specificity through sugar inhibition studies. Recombinant LecB 3K, along with some other recombinantly produced prokaryotic lectins, was used in unison with the other eukaryotic lectins to glyco-profile the cells. In this chapter the lectins that show significant differences in cell surface binding are highlighted and discussed. The binding of HPA for both SW480 and SW620 proved to be very interesting and along with the data generated in the previous chapters, the HPA binding will be compared and contrasted with current literature.

5.2 Optimising the parameters for non-fixed cell probing and sample preparation of SW480 and SW620 for flow cytometry

Full methodological details for this section are provided in Section 2.28. The gating strategy outlined in section 2.28.4 was used for all subsequent experiments using SW480 and SW620. The forward scatter area (FSC-A) and side scatter area (SSC-A) for SW620 and SW480 are different based on the cells being different in size. For this reason all gating carried out was cell line specific and the live gates were determined using unstained populations for each cell type (See Section 2.28). Cells were cultured as per section 2.25. The SW480 cells grown to 80% confluency after 48 hours routinely produced between 5-7 million cells per flask at approximately 98-99% cell viability. The SW620 cells grown to 80% confluency after 48 hours routinely produced between 9-11 million cells per flask at approximately 98-99% cell viability as well. The methods outlined in section 2.28 were dictated based on various experiments carried out to identify the correct parameters. Due to the two cell types being adherent cells and needing to be trypsinised, washed, and probed all within the shortest time scale available, experimental parameters needed to be defined and optimised. Literature suggests probing with lectins at a concentration of between 1-10 µg/mL, while incubating the cells in the dark at 25 °C or 4 °C for 30-60 minutes (Kekalainen et al. 2015 and Batisse et al. 2004). The variance in the cell incubation time, amount of lectin used and temperature of incubation time suggest a variability based on the type of cell and the specific lectins used in each experiment. To limit this variability experiments were carried out to find the optimum; cell number, lectin concentration used to probe and growth time of cells after seeding. The doubling time for SW620 is approximately 26 hours, whereas the doubling time for SW480 is approximately 38 hours (American Type Culture Collection, 2012). To ensure the cells have attached to the plate and gone through one full growth phase, the cells are seeded and left to grow at 37 °C for 24, 48 and 72 hours then probed with a small panel of six lectins to determine binding efficiency. The panel of lectins used were; Con A, DSL, ECL, GNL, Jacalin and PNA. The specificity for these lectins is shown in section 2.24. These lectins bind to a variety

of different cell surface glycans and show significant binding to both SW480 and SW620. Due to the difference in doubling time between the two cell types, the original seeding of the cells was adjusted to try and normalise the number of cells grown for each time point. In figure 5.1 the different time points are shown to highlight specific binding of the cells after each growth time point.

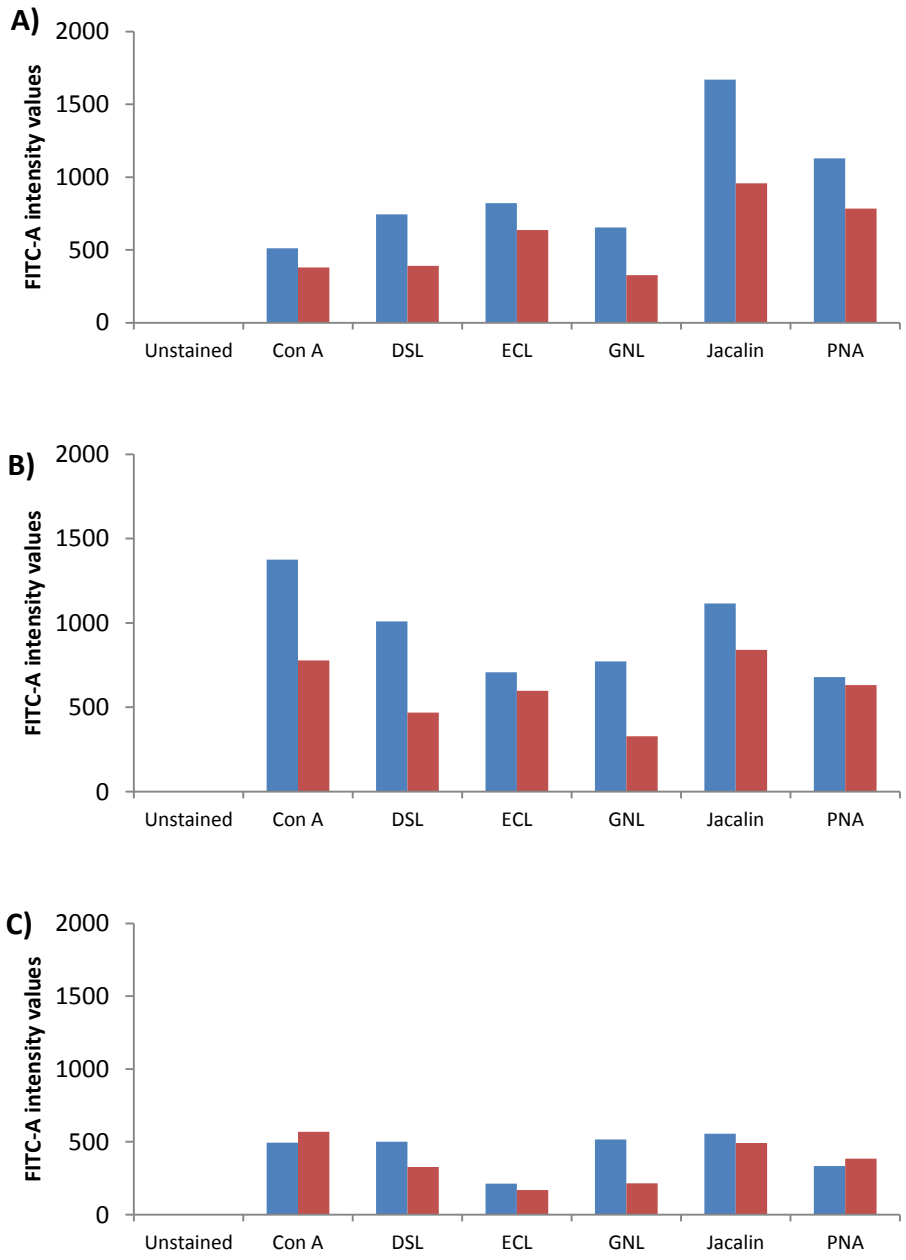


Figure 5.1: Lectin probing of SW480 and SW620 cells over 3 growth time points of 24, 48 and 72 hours. Cells were grown for the specified times and probed with the same concentration of lectin for the same duration of time and samples were analysed using flow cytometry. A) Growth time 24 hours, B) Growth time 48 hours, C) Growth time 72 hours. (n=1)

SW480 and SW620 cells were originally epithelial cells and behave as such in culture. The cells need to be attached to each other in order to ensure correct cellular function. After 24 hours the cells were not all attached and not given enough time to ensure a full growth cycle had taken place. After 48 hours the cells had undergone at least one full growth cycle and the cells still maintained a high level of cell viability. The mean fluorescent intensity values for both 24 and 48 hours were considerably higher than that of the samples tested after 72 hours. 48 hours was decided as the optimal growth time for cells before probing with lectins and sample application on the flow cytometer.

After the growth time analysis was carried out, the optimal amount of cells and concentration of lectin per sample was investigated. A large number of cells can be cultivated in a relatively short period of time so the optimal number of cells probed with lectin is believed to be between 100,000 and 1 million cells. This study carried out lectin probing at four different concentrations of cells ranging from 125,000 to 1,000,000 cells. The study was carried out using 5 µg/mL of the lectin Con A which binds mammalian N-glycans. To determine the correct numbers of cells to use per sample three parameters were used, these were; percentage of cells in the live gate, mean fluorescent intensity and the ratio of mean fluorescent intensity between the two cell lines. In table 5.1 these values are highlighted.

Cell number	% of cells in live gate		Mean FITC intensity		FITC Ratio (SW480:SW620)
	SW480	SW620	SW480	SW620	
1,000,000	52.8	45.8	1257	742	**1.69:1**
500,000	50.9	42.2	2551	1659	**1.53:1**
250,000	47.8	45.4	4496	2678	**1.67:1**
125,000	48.2	46	7148	5702	**1.25:1**
Unstained	82.2	81.8	3.67	0.79	**N/A**

Table 5.1: Lectin probing of SW480 and SW620 cells changing the number of cells probed. Four different amounts of cells were probed with the same concentration of the lectin Con A. Parameters of the experiment are cell number, percentage of cells in the live gate, mean fluorescent intensity and the ratio of FITC intensity between the two cell types.

The lower the number of cells the higher fluorescent intensity observed, this is most likely caused by over saturation of lectin. Using a lower number of cells, 125,000, proved difficult due to the loss of cells at each centrifugation step and the cell death attributed to lectin binding. Running 10-15,000 cells on the flow cytometer per sample took longer and therefore increased the time between the first and last tube of the experiment being run on the flow cytometer. With non-fixed cell staining the shortest possible time between the first and last sample being run is optimal. For these reasons the cell number of 250,000-500,000 was determined to be the optimal for cell surface probing with lectins.

The saturation point of each cell type of each lectin is subjective and can change for each lectin. Some lectins cause a reduction in cell viability, as was stated in section 2.28.4, and probing at higher concentrations causes an increase in the amount of cell debris and overall number of non-viable cells. Four lectins highlighted as cytotoxic and non-cytotoxic were used to determine an appropriate lectin concentration to probe these cells with. The four lectins chosen were AAL, PNA, SNA and NPL. Binding specificities for these four lectins are noted in Section 2.24. When the concentration of NPL and PNA are increased the number of cells within the 'live' gate remains relatively constant whereas when the concentration of AAL and SNA are increased the number of cells in the 'live' gate significantly diminishes. In table 5.2 the percentage of cells within the live gate, as well as the relative fluorescent intensities observed for varying concentrations of lectin are noted. Due to the decreased number of viable cells for flow cytometric analysis, the optimal concentration of lectin chosen for the remainder of the project was 5 µg/mL.

Concentration of lectin (µg/mL)	Percentage of cells within the live gate		FITC intensity	
	SW480	SW620	SW480	SW620
SNA 1	80.7%	70.7 %	89.4	33.6
SNA 5	77.1%	75.0 %	2327	2576
SNA 10	45.7%	31.1 %	2381	2977
SNA 20	33.1%	29.6 %	3099	2782
SNA 50	30.9%	27.0 %	2758	2617
NPL 1	83.0 %	83.8 %	34.3	28.1
NPL 5	79.1 %	83.2 %	459	123
NPL 10	78.3 %	85.1 %	506	170
NPL 20	77.8 %	85.3 %	600	386
NPL 50	77.5 %	81.6 %	538	405
AAL 1	66.3%	57.1%	397	323
AAL 5	47.1%	53.5%	1452	800
AAL 10	23%	23.8%	1593	1715.5
AAL 20	23.5%	21.5%	1579	1316
AAL 50	23.5%	21.0%	2459	1234
PNA 1	75.7%	79.3%	46.9	7.905
PNA 5	77.4%	77.3%	789.5	1037
PNA 10	73.9%	79.0%	768.5	807
PNA 20	77.0%	81.6%	655.5	917
PNA 50	74.9%	78.1%	351.5	1197

Table 5.2: SW480 and SW620 cells probed using different concentrations of four lectins. The four lectins used are SNA, NPL, AAL and PNA at a concentration range of 1-50 µg/mL.

5.3 Flow cytometric analysis of SW480 and SW620 using recombinant prokaryotic lectins and commercially purchased eukaryotic lectins

The parameters determined in Section 5.2 were applied to all the subsequent flow cytometry experiments and a panel of 23 lectins (3 prokaryotic and 20 eukaryotic) was used to identify glycoprofile differences and similarities between SW480 and SW620. The lectins used in this section are shown below in table 5.3. The lectins were chosen due to their diverse binding affinities (Section 2.24). The most common sugars found in complex mammalian N- and O- linked glycans were accounted for, as well as determining specific glycan linkages.

Lectins used as probes for flow cytometry		
AAL	Con A	DBA
DSL	ECL	GNL
GSL I	GSL II	HPA
Jacalin	LCA	MAL I
MAL II	NPL	PNA
RCA	SBA	SNA
UEA I	WGA	
Recombinant Lectins used as probes for flow cytometry		
AAL 2	LecA	LecB 3K

Table 5.3: Eukaryotic and Prokaryotic lectins used as probes for flow cytometry.

The SW480 and SW620 cell lines used are significantly different in cell size, shown by number of available cells per plate at 80-100% confluency as well as the FSC-A and SSC-A of unstained populations monitored by the flow cytometer. FSC-A is generally used as a parameter to determine relative cell size. It is debated in the literature the exact size parameter it measures, i.e. is it cell surface area, cell volume or the diameter of the cell, all relative to the calibration beads used to ensure the lasers are correctly aligned (Taylor 2016; Knijnenburg et al. 2011; Biosciences 2000). Cells trypsinised and re-suspended in buffer for analysis on the flow cytometer lead to the cells being in a relatively spherical state. The SW480 cells are significantly larger than SW620 cells, which can be seen by the difference in FSC-A values shown in figure 5.2. The difference in relative size of each cell type signifies a potential surface area change between the two cells. This surface area change allows for the additional binding of each lectin as a percentage of cell surface area. For this reason the absolute values for binding being higher in general for SW480 cells over SW620 cells do not necessarily represent an increase in that specific glycan target. Some of the lectins that bind with a ratio close to 2:1 for SW480:SW620 are believed to be representing the lectins that do not highlight a significant difference in glycan per surface area on either cell type. Again for this reason the lectins binding SW480 above this relative 2:1 ratio are believed to be showing an increase in the presence of their glycan target on the surface of SW480. The inverse is also used to signify the increase in the presence of specific glycan targets on the SW620 cell surface. These are reflected by the lectins that bind close to a 1:1 ratio for between the 2 cell lines. The three groups; ‘normal’ ratio binding, increased glycan targets on the surface of SW480 and increased glycan targets on the surface of SW620 are represented in figures 5.3, 5.4 and 5.5, respectively.

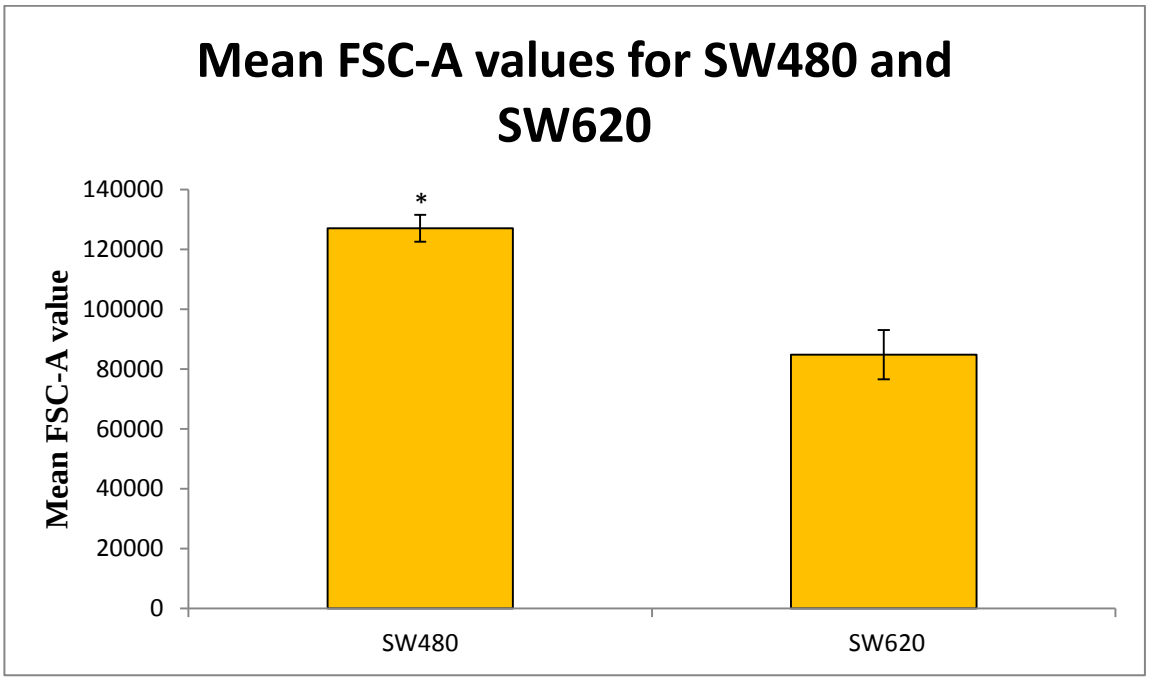


Figure 5.2: The mean FSC-A values for SW480 and SW620 cells viewed on a flow cytometer to determine relative cell size. Data are means ± standard deviations. * represents $P < 0.05$, n=3.

A smaller number of lectins, three out of the panel, used showed significant increases in lectin binding on SW480 over SW620 whereas another small number, five out of the panel, of lectins showed increases in lectin binding on SW620 over SW480. Figures 5.3, 5.4 and 5.5 show the differences in binding exhibited for different lectins on SW480 and SW620. All binding was expressed as a ratio of binding for SW480 compared to SW620. The relative fluorescent intensities recorded for each lectin on the SW620 cells are normalised to 1 and the fluorescent intensities recorded for SW480 are expressed as a ratio relative to that. Each lectin exhibits different mean fluorescent intensities due to the differences in each lectin. The potential binding of the fluorophores used to each lectin can change and each individual lectin is not compared to another lectin. Some lectins that did not bind were commonly used as negative controls, such as GSL I and GSL II. Both GSL I and GSL II showed no binding and will not be represented on any of the subsequent graphs. This is shown in figure 5.6 that compares the positive cells in

the population to the negative cells in the population. Another lectin that shows both relatively strong binding as well as no-binding within the cell population is PNA. PNA seems to bind anywhere from 40-60% of the cells in a given population, showing the heterogeneity in the cell lines. Due to both DBA and PNA having large numbers of FITC-A positive and negative cells within the live gate the mean fluorescent intensity for each population was not used to determine the ratio of fluorescent intensity between the two cell lines. WGA findings were mixed with the ratio of binding between SW480 and SW620 cells ranging from 1:1 to 2:1. This was caused by WGA's cytotoxic effects leading to low numbers of cells within the live gates. WGA is not used in figures 5.3, 5.4 or 5.5 but its histograms are shown in figure 5.6.

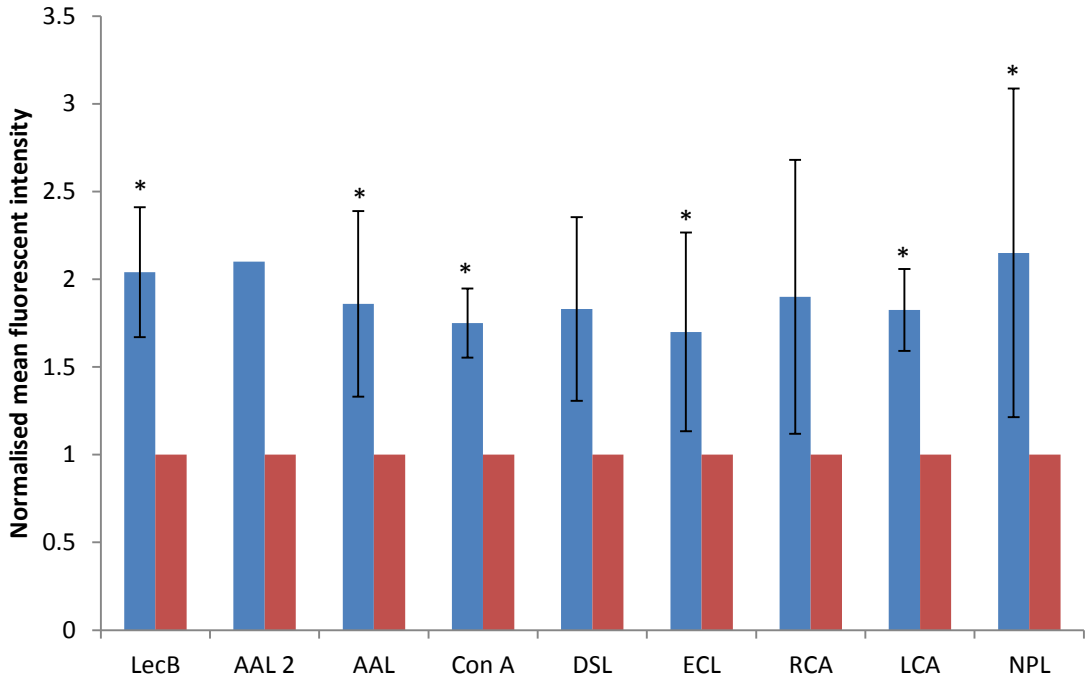


Figure 5.3: Lectin binding to SW480 and SW620 cells, showing a relative 2:1 binding pattern for SW480:SW620. These nine lectins show a normal binding pattern for SW480 and SW620 due to the difference in cell sizes. SW620 cells have been normalised to 1. SW480 cells are represented by blue histograms, while SW620 cells are represented by red histograms. Data are means ± standard deviations. * represents $P < 0.05$, n=3. (AAL 2 experiment only carried out once).

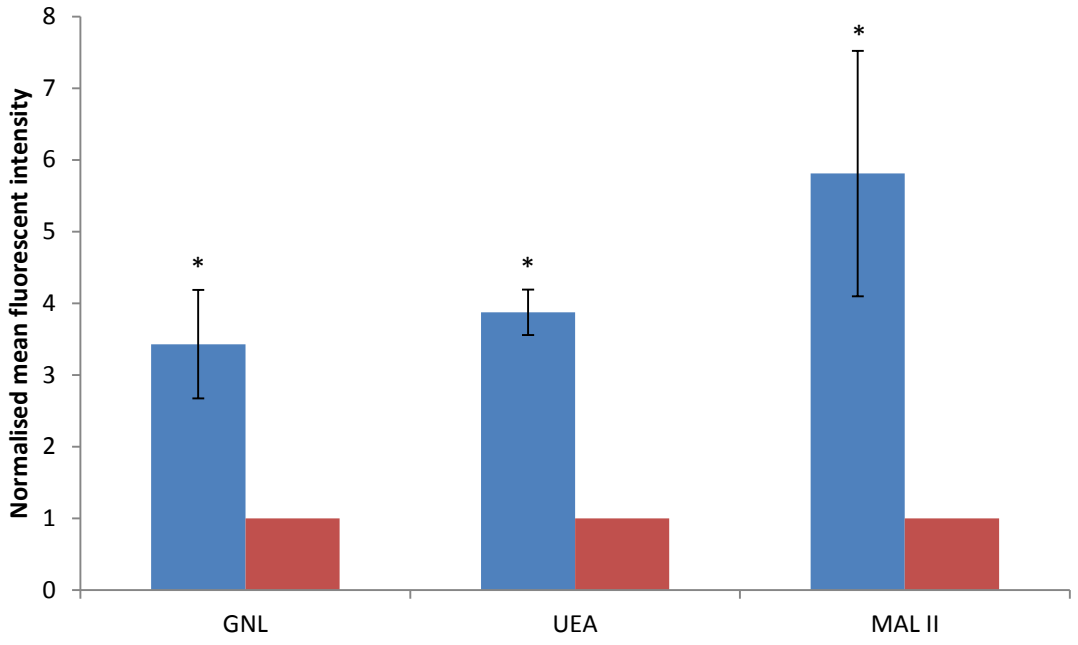


Figure 5.4: Lectin binding to SW480 and SW620 cells, showing an increase in binding for SW480 over SW620. These three lectins bind more to SW480 than SW620. SW620 cells have been normalised to 1. SW480 cells are represented by blue histograms, while SW620 cells are represented by red histograms. Data are means ± standard deviations. * represents $P < 0.05$, n=3.

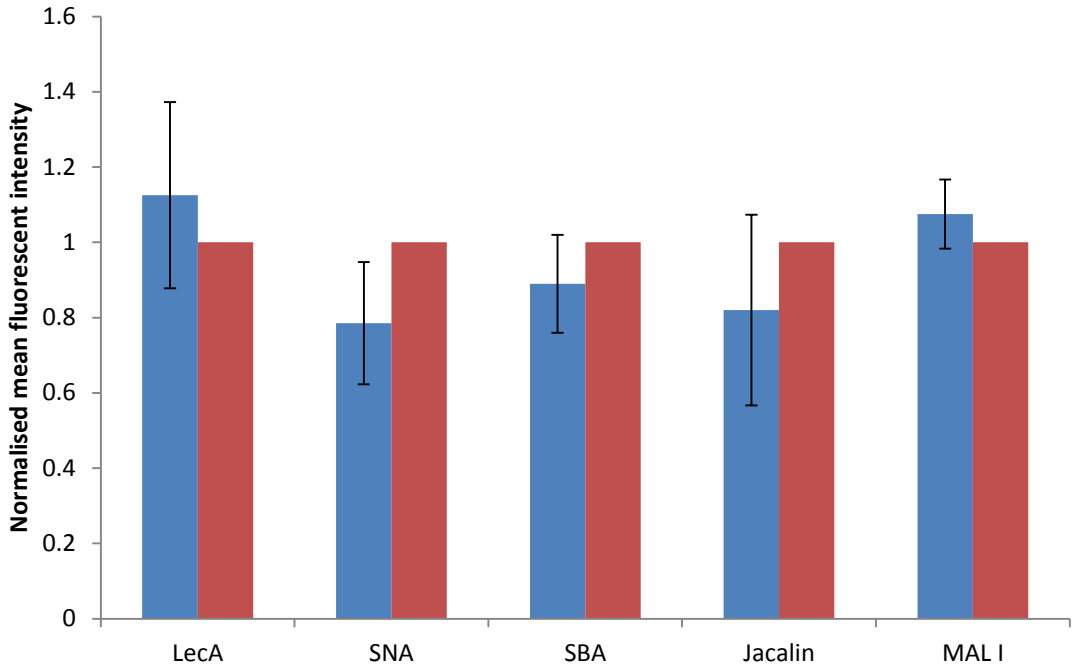
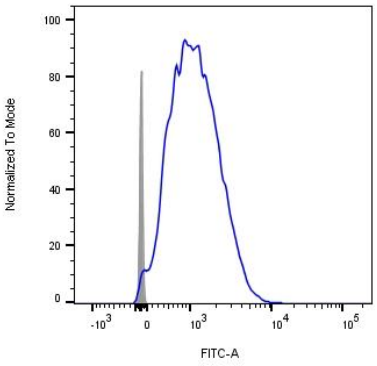
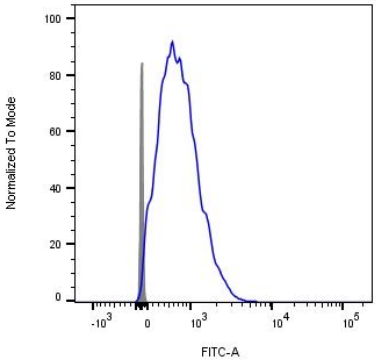
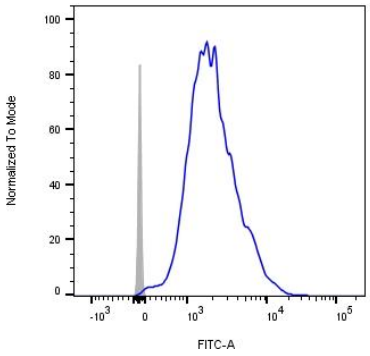
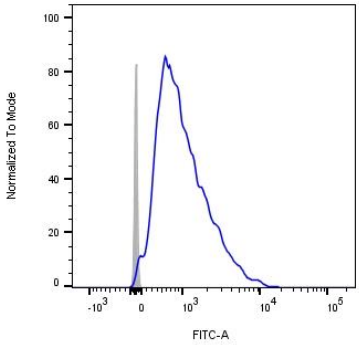
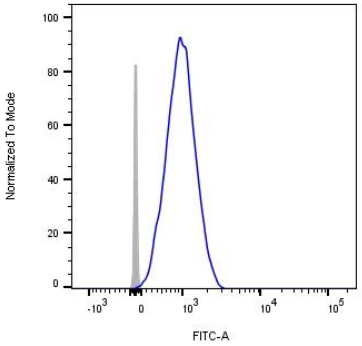
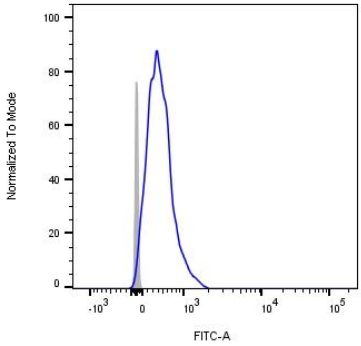
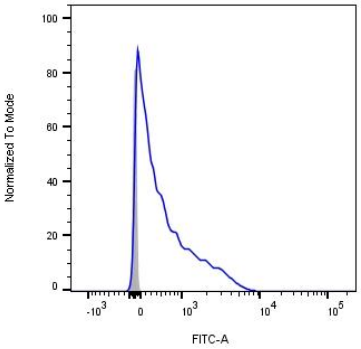
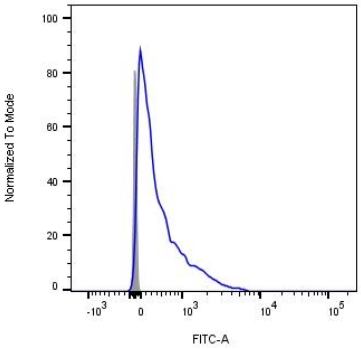


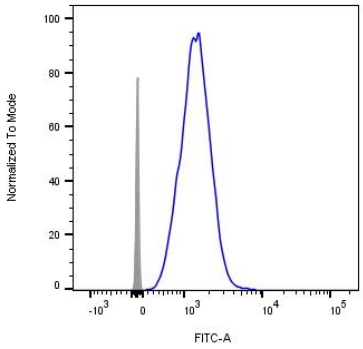
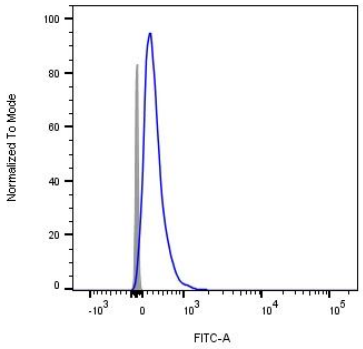
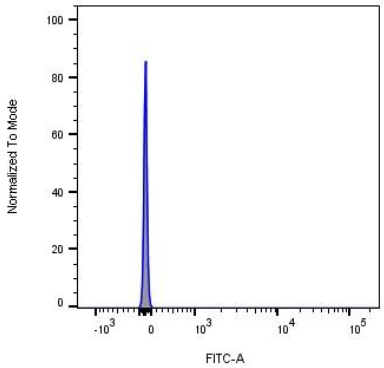
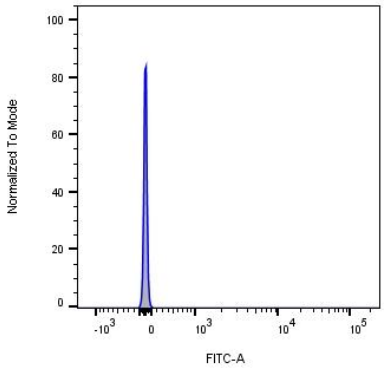
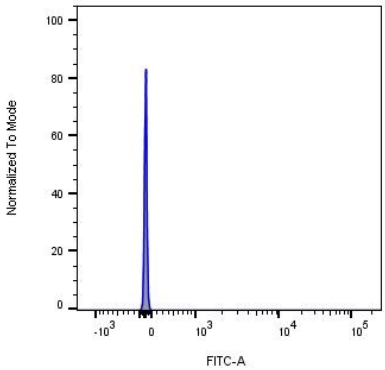
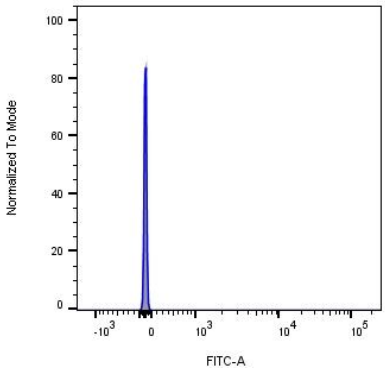
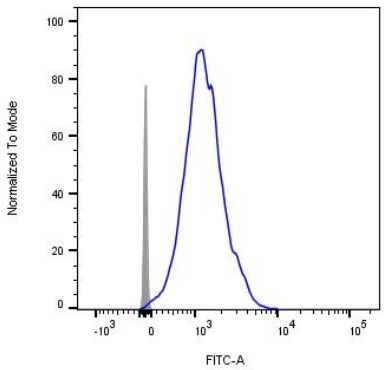
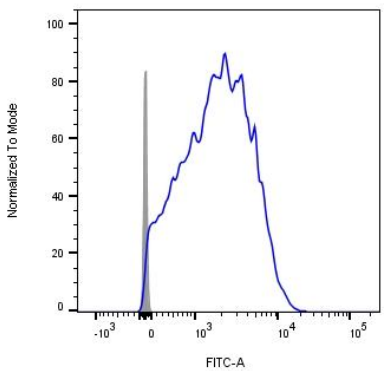
Figure 5.5: Lectin binding to SW480 and SW620 cells, showing an increased binding to SW620 over SW480. Similar binding infers increased binding to SW620 due to decreased cell size when compared to SW480. SW620 cells have been normalised to 1. SW480 cells are represented by blue histograms, while SW620 cells are represented by red histograms. Data are means ± standard deviations. * represents $P < 0.05$, n=3.

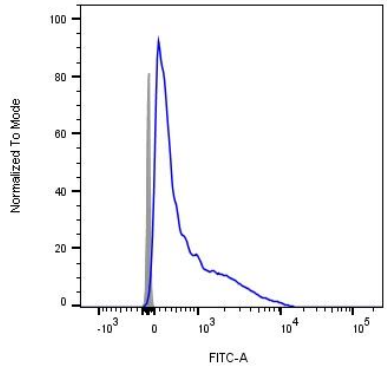
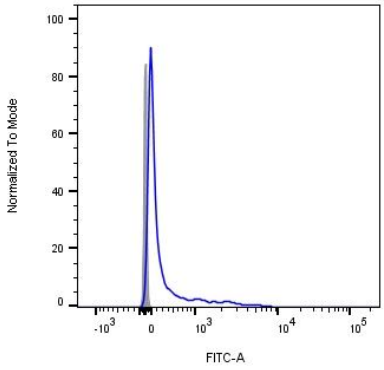
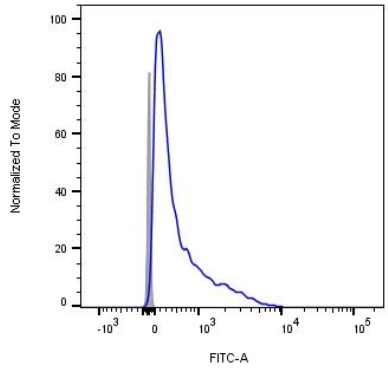
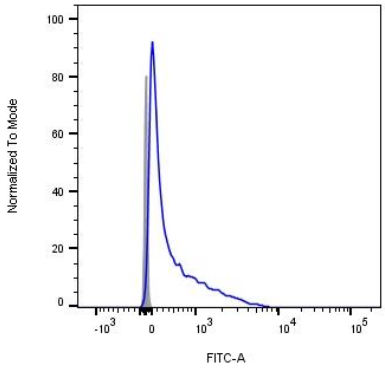
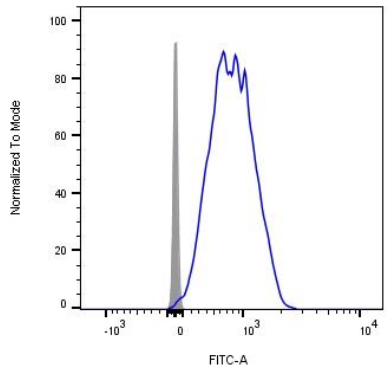
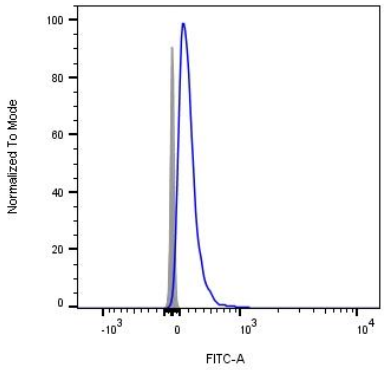
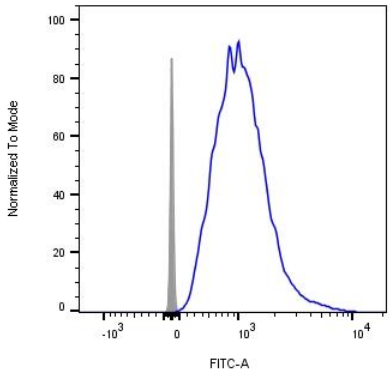
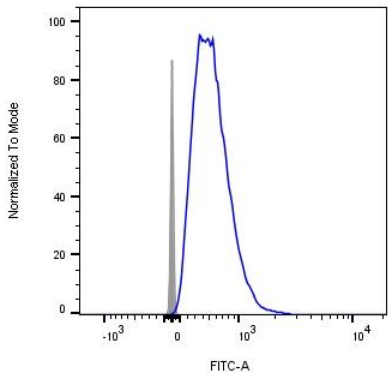
5.3.1 Histogram analysis of lectin binding to SW480 and SW620

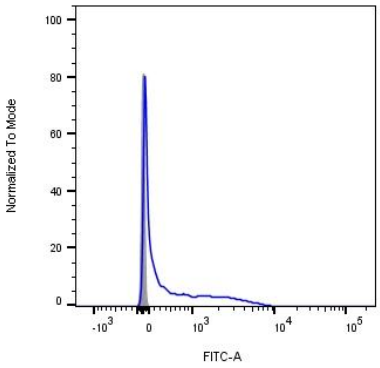
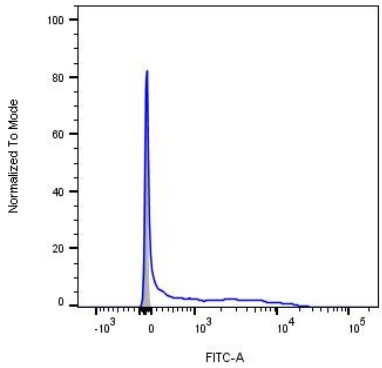
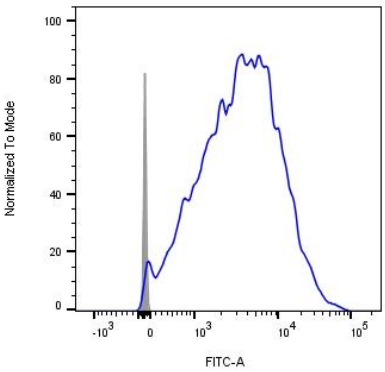
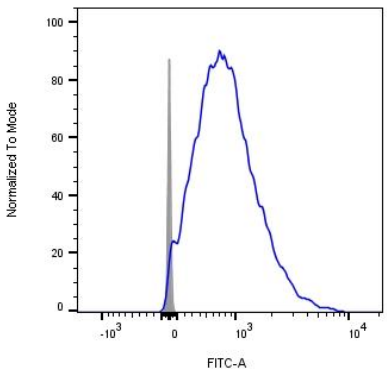
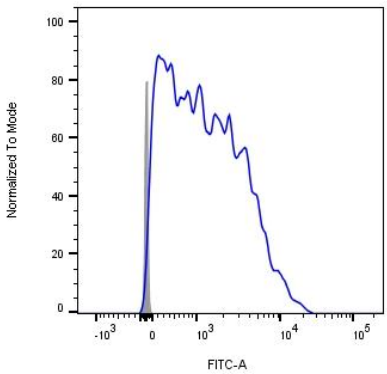
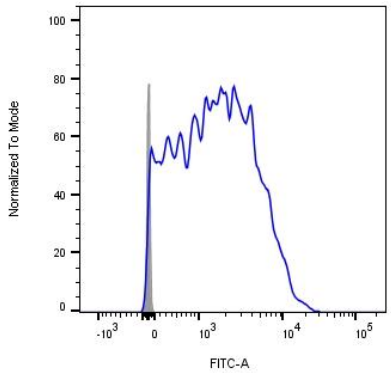
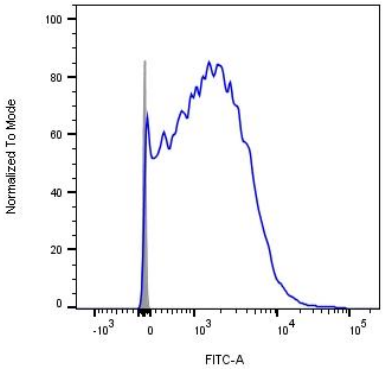
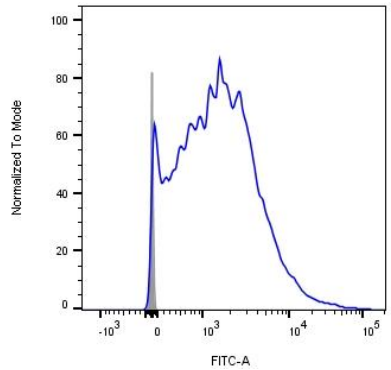
All data in figures 5.3 to 5.5 were acquired using the average of three mean fluorescent intensity values from different repeated experiments. These experiments allowed for the mean fluorescent intensity (FITC-A) values to be obtained using FlowJo, see section 2.28. FlowJo generates histograms that allow for similarities and differences between the binding profiles of each lectin on the two cell lines to be compared. Some of the lectins bind relatively uniformly giving a sharper peak on a histogram whereas others give a much broader peak. For these reasons the error bars on figures 5.3 to 5.5 are greater than desired. The histogram figures below allow for direct comparison between the two cell lines for each individual lectin. The unique differences, such as binding specificity and available biotin labels, between each lectin means a direct comparison between lectins themselves is more difficult to control for.

Figures 5.6 and 5.7 show the histograms for all the eukaryotic and prokaryotic lectins, excluding DBA and HPA, with the FITC-A compared to the unstained control cell population. Each graph is normalised to the mode to allow for clear histogram images. The histogram peak distribution highlights the differential binding of each lectin to that specific cell line. Some lectins have a smaller number of cells that have a high level of lectin binding. This causes a tailing effect in the histogram that shows the fluorescent intensity for these cells is considerably higher than the average in the population. The tailing effect is represented by the histograms for LCA, MAL I, PNA and ECL to a large extent. The amount of positive and negative cells in the live gate remained relatively constant for LCA, MAL I and ECL, but PNA probing varied from between 40-70% for relative fluorescent intensity (FITC-A).

	SW480	SW620
AAL		
Con A		
DSL		
ECL		

	SW480	SW620
GNL		
GSL I		
GSL II		
Jacalin		

	SW480	SW620
LCA		
MAL I		
MAL II		
NPL		

	SW480	SW620
PNA		
RCA		
SBA		
SNA		

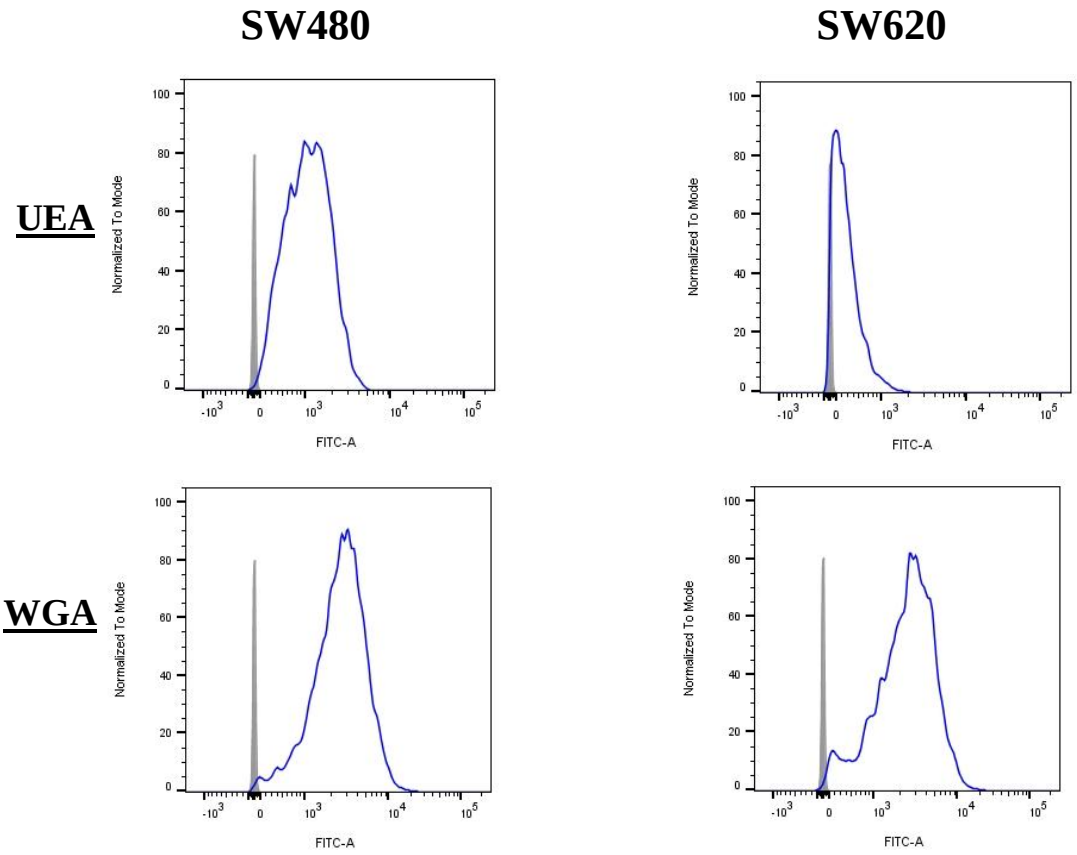


Figure 5.6: Analysis of eukaryotic lectins binding to both SW480 and SW620 cells as represented by histograms. The histograms show Lectin probed cells as represented by non-filled dark blue histograms and the unstained cell population, not probed with lectin, are represented by filled grey histograms. The lectin name is to the left of each histogram. The histograms were both normalised to the mode, n = 3. Each lectin has a different specificity (Section 2.24). All lectins used were purchased from Vector Labs®. The SW480 probed cells are on the left hand side, while the SW620 probed cells are on the right hand side.

The recombinant lectins used were all biotinylated (Section 2.22) so that the streptavidin linked fluorescent dyes could be used for both the recombinant lectins and the eukaryotic lectins. The three recombinant lectins were AAL 2, LecA and LecB 3K. The AAL 2 and LecA lectins were also recombinantly produced and purified. Figure 5.7 shows the three histograms of these lectins compared to the unstained negative control.

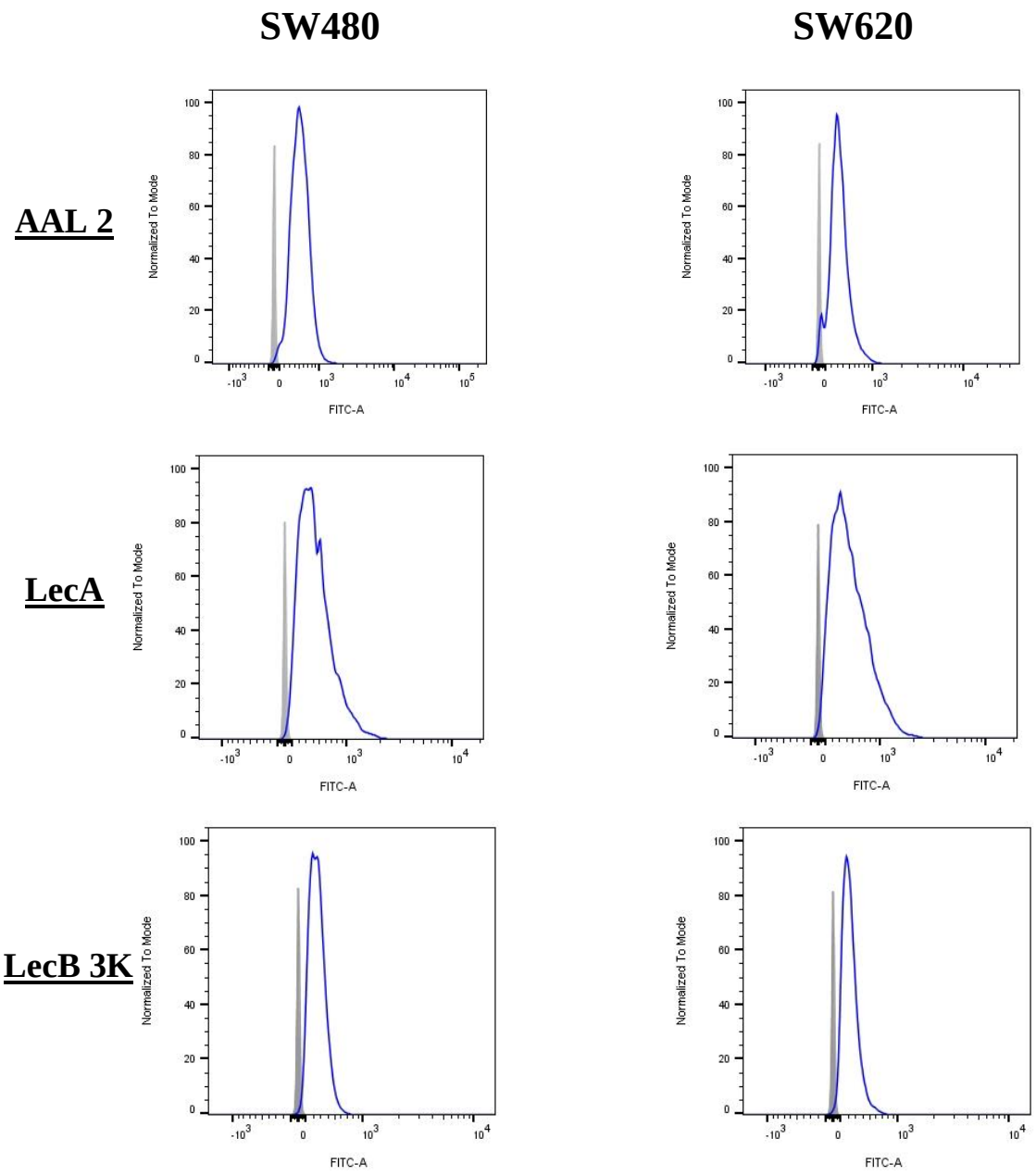


Figure 5.7: Analysis of recombinant prokaryotic lectins binding to both SW480 and SW620 cells. The histograms show Lectin probed cells as represented by non-filled dark blue histograms and the unstained cell population are represented by filled grey histogram. The lectin name is to the left of each histogram. The histograms are normalised to the mode, n = 3.

DBA is another lectin that highlighted heterogeneity, similar to LCA, MAL I and PNA, but to a much lesser degree. DBA for SW620 shows 4-12% positively stained cells within the live gate whereas DBA for SW480 shows 1-3% positively stained cells within the live gate. Due to these cells being largely negative, the positive population was compared to the negative population in histogram analysis. In figure 5.8 the DBA lectin comparisons to its unstained negative control cannot show the significantly high binding targets, whereas in figure 5.9 the FITC-A positively stained cells compared to the FITC-A negative cells show the extremely positive cells.

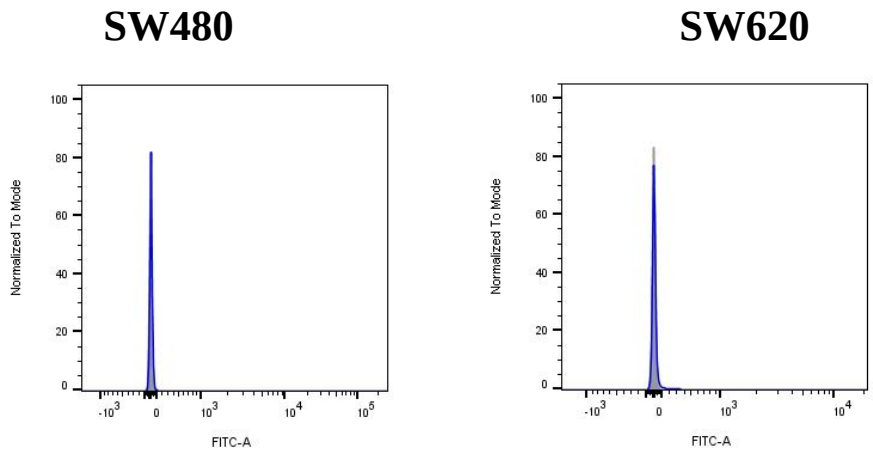


Figure 5.8: Analysis of DBA probing to both SW480 and SW620 cells. The histograms show DBA probed cells as represented by non-filled dark blue histograms and the unstained cell population is represented by filled grey histograms. The histograms have been normalised to mode, n = 3.

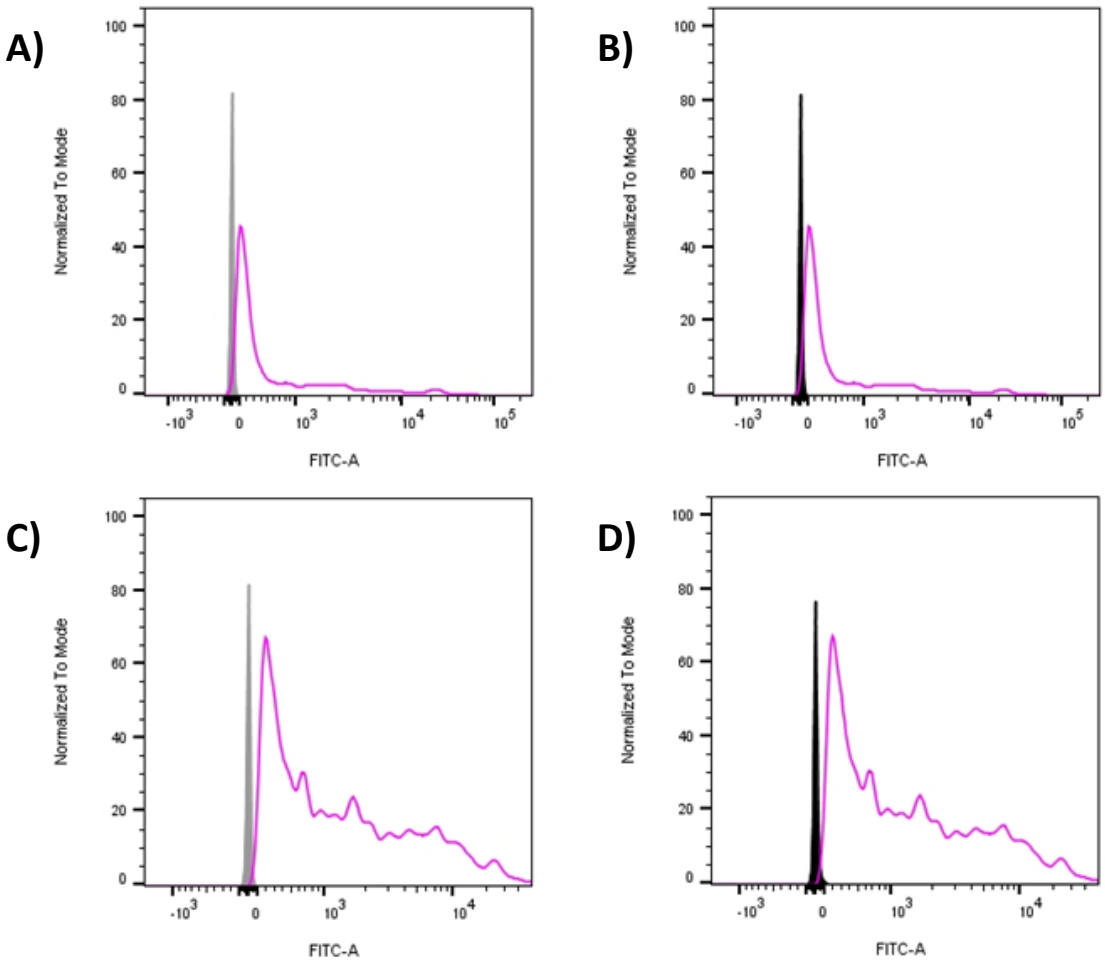


Figure 5.9: Analysis of DBA positively stained cells compared to both unstained cells and the DBA negative cells. A) SW480 unstained cells overlaid with FITC-A positive cells, B) SW480 FITC-A negative cells overlaid with FITC-A positive cells, C) SW620 unstained cells overlaid with FITC-A positive cells, D) SW620 FITC-A negative cells overlaid with FITC-A positive cells. Grey represents unstained cells, Black represents FITC-A negative cells probed with DBA and Pink represents FITC-A positive cells probed with DBA. All histograms were normalised to mode, n = 3.

5.4 Lectin binding inhibited using competitive free sugars tested using flow cytometry

In Chapter 4 a number of lectins were pre-incubated with different concentrations of their corresponding free sugar. These free sugars caused inhibition in lectin binding to the surface of the cells. Using sugars to competitively inhibit binding signifies the specificity of the lectin binding and proves the lectins are binding via their sugar binding site and not through any other protein-protein interactions. To ensure the inhibition is specific each lectin chosen to be inhibited is subjected to an inhibitory sugar, as well as a sugar the lectin has no affinity for that will not greatly affect lectin binding, i.e. non-inhibitory. For the same reason as in chapter 4 some lectins, WGA, SNA and MAL II, were not chosen for sugar inhibition due to the pH change in buffer with addition of sugar or where there are additional wash steps with acid or basic solutions. Out of the twenty three lectins used for glyco-profiling, four eukaryotic lectins (AAL, ECL, Jacalin and NPL) and two recombinant prokaryotic lectins (LecA and LecB) were chosen to be tested with two different inhibitory and non-inhibitory sugars. Some other lectins were tested in this way and those histograms can be seen in appendix B.1. Table 5.4 shows the lectins chosen as well as the two different sugars used. As in Chapter 3 and Chapter 4, not all lectins subjected to a high level of free sugar can be specifically inhibited. This is caused by the lectins need to bind complex glycans that are exhibiting specific linkages with the under lying glycan structure.

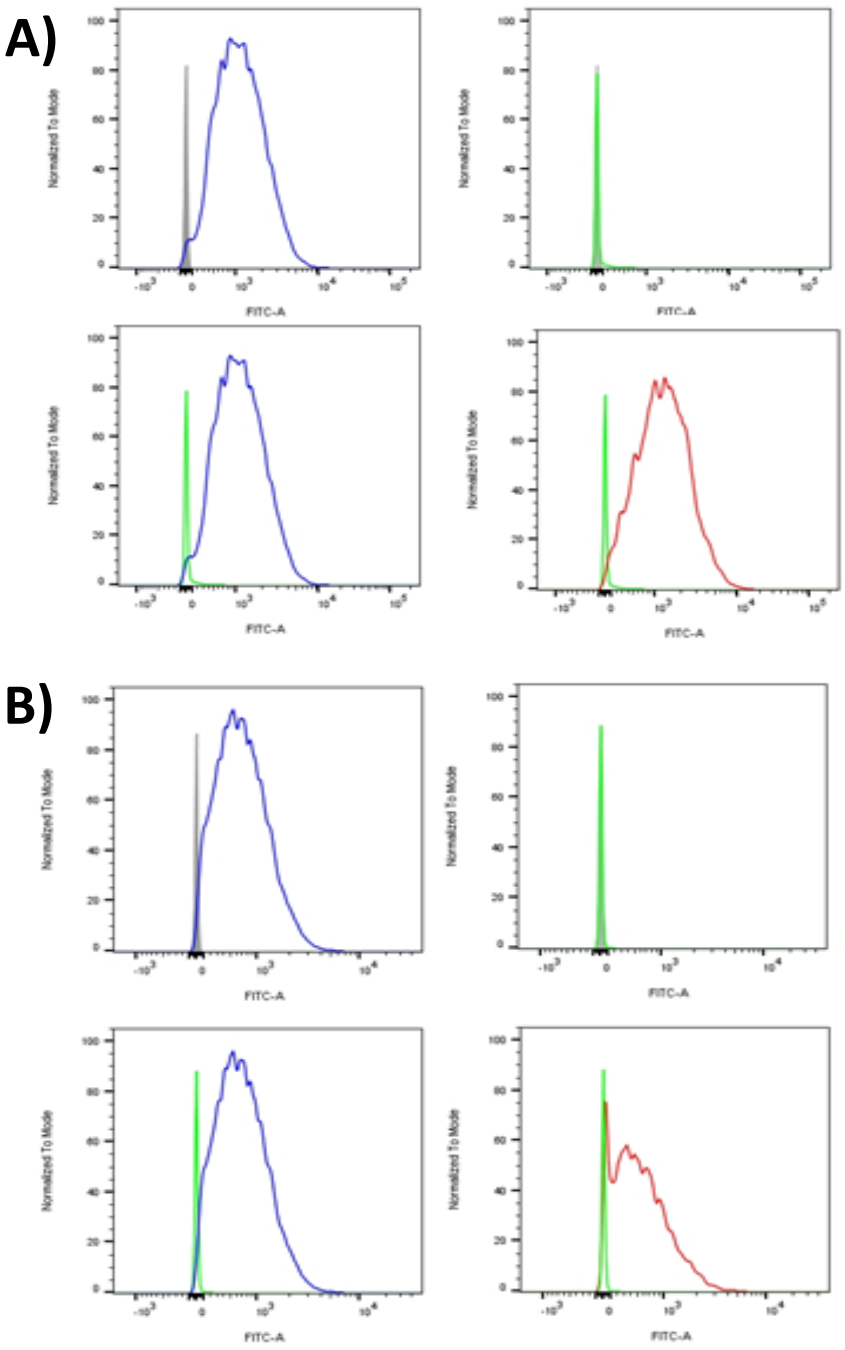
Each lectin was represented by a quad-panel of histograms with 2 conditions represented in each panel. Panel 1 shows unstained cells overlaid with lectin stained cells; panel 2 shows unstained cells overlaid with lectin stained cells whereby the lectin was first pre-incubated with its corresponding inhibitory sugar; panel 3 shows lectin stained cells whereby the lectin was first pre-incubated with its corresponding inhibitory sugar overlaid with cells probed with lectin and panel 4 shows lectin pre-incubated with its corresponding inhibitory sugar overlaid with lectin pre-incubated with a non-inhibitory sugar. These panels highlight the specificity of each lectin by showing the minimal

amount of inhibition observed when pre-incubating a lectin with a sugar it has no affinity for when compared to near total inhibition when pre-incubated with a sugar the lectin has affinity for.

Lectin sample	Inhibitory sugar	Non-inhibitory sugar
AAL	200 mM L-fucose	200 mM D-fucose
ECL	200 mM lactose	200 mM L-fucose
Jacalin	100 mM melibiose	100 mM glucose
NPL	200 mM α-methylmannoside / α-methylglucoside	200 mM glucose
LecA	200 mM galactose	200 mM L-fucose
LecB 3K	200 mM L-fucose	200 mM D-fucose

Table 5.4: Panel of lectins pre-incubated with inhibitory and non-inhibitory free sugars used to probe SW480 and SW620 cells. The inhibitory sugars and their concentrations are supplied by Vector Laboratories®. The concentration of non-specific sugar was determined based on the concentration suggested for inhibition.

The majority of specific competitive inhibitory sugars caused lectin binding to be nearly completely inhibited though although there are some lectins that cannot be fully inhibited, such as jacalin. In figure 5.12 the green histograms that show jacalin pre-incubated with melibiose only showed partial inhibition where there was still a significant amount of the cell population in the live gate fluorescently labelled. Lectin specificity is determined by the probes ability to bind to the cells in the presence of a non-competing sugar as well as its inability to bind to the cells in the presence of a competing sugar.



Colour scheme for sugar inhibition histograms;

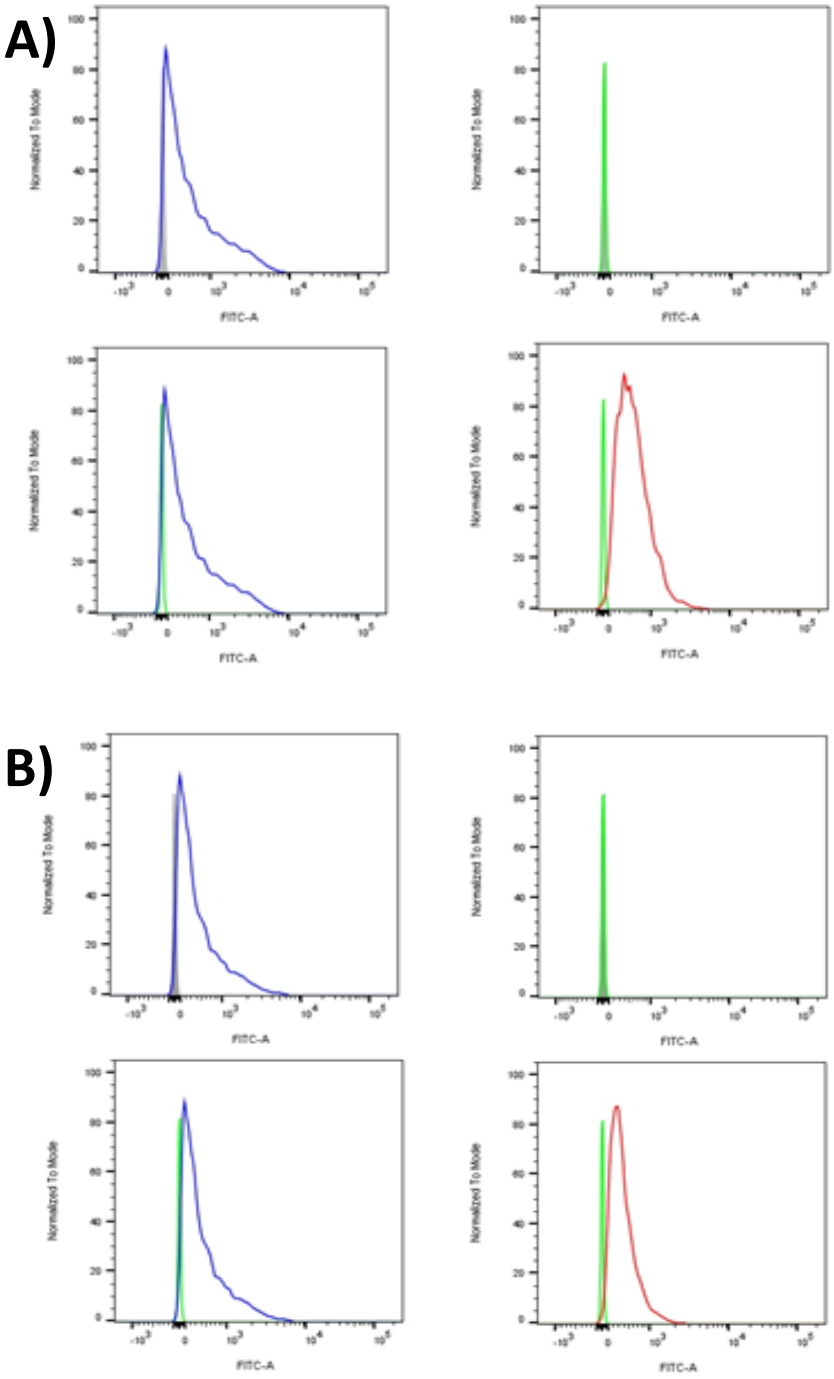
Grey = Unstained Cells

Blue = Cells probed with lectin

Green = Cells probed with lectin pre-incubated with inhibitory sugar

Red = Cells probed with lectin pre-incubated with non-inhibitory sugar

Figure 5.10: SW480 and SW620 cells probed with AAL and AAL pre-incubated with inhibitory and non-inhibitory sugars. A) SW480 cells, B) SW620 cells. Inhibitory and non-inhibitory free sugars are shown in table 5.4. Histogram colour scheme displayed in text box adjacent to A) and B).



Colour scheme for sugar inhibition histograms;

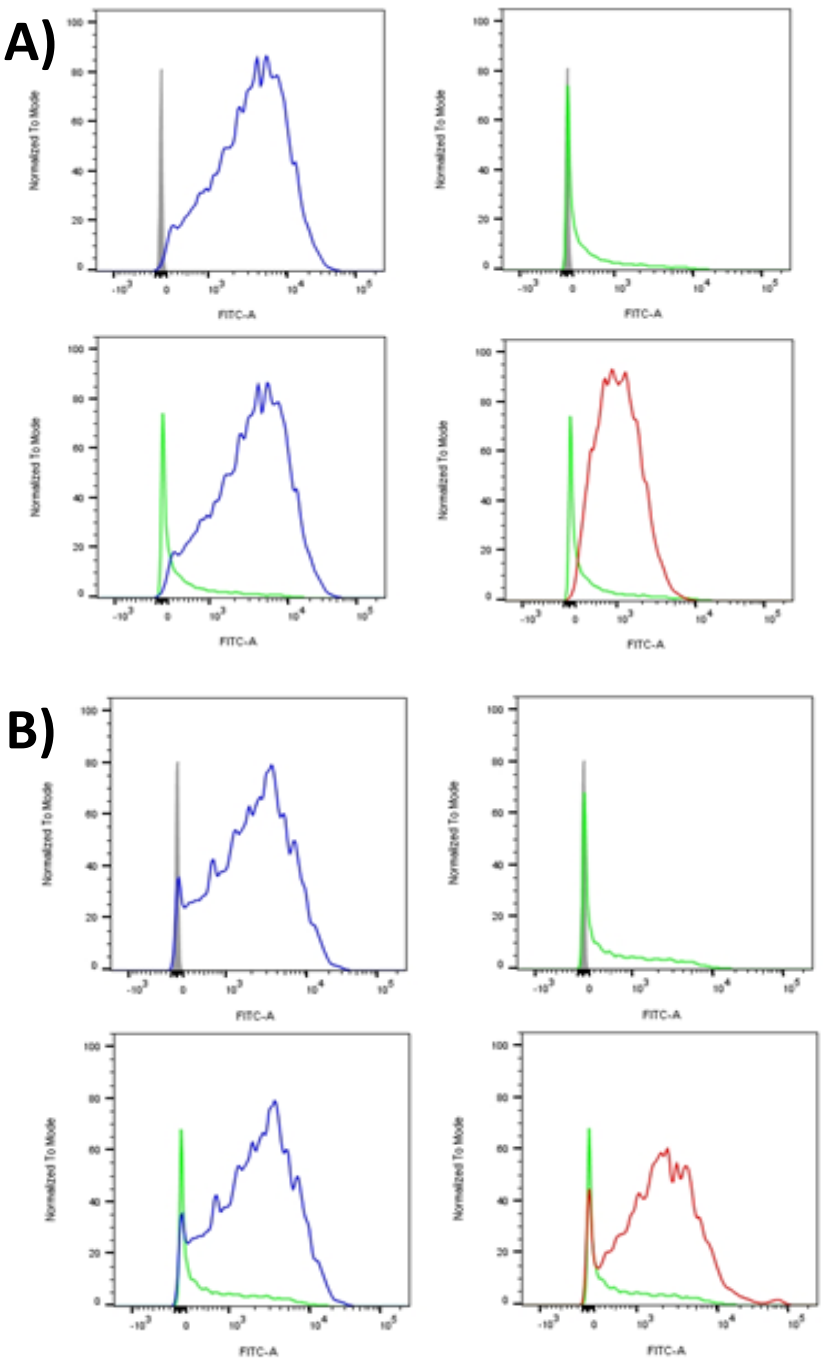
Grey = Unstained Cells

Blue = Cells probed with lectin

Green = Cells probed with lectin pre-incubated with inhibitory sugar

Red = Cells probed with lectin pre-incubated with non-inhibitory sugar

Figure 5.11: SW480 and SW620 cells probed with ECL and ECL pre-incubated with inhibitory and non-inhibitory sugars. A) SW480 cells, B) SW620 cells. Inhibitory and non-inhibitory free sugars are shown in table 5.4. Histogram colour scheme displayed in text box adjacent to A) and B).



Colour scheme for sugar inhibition histograms;

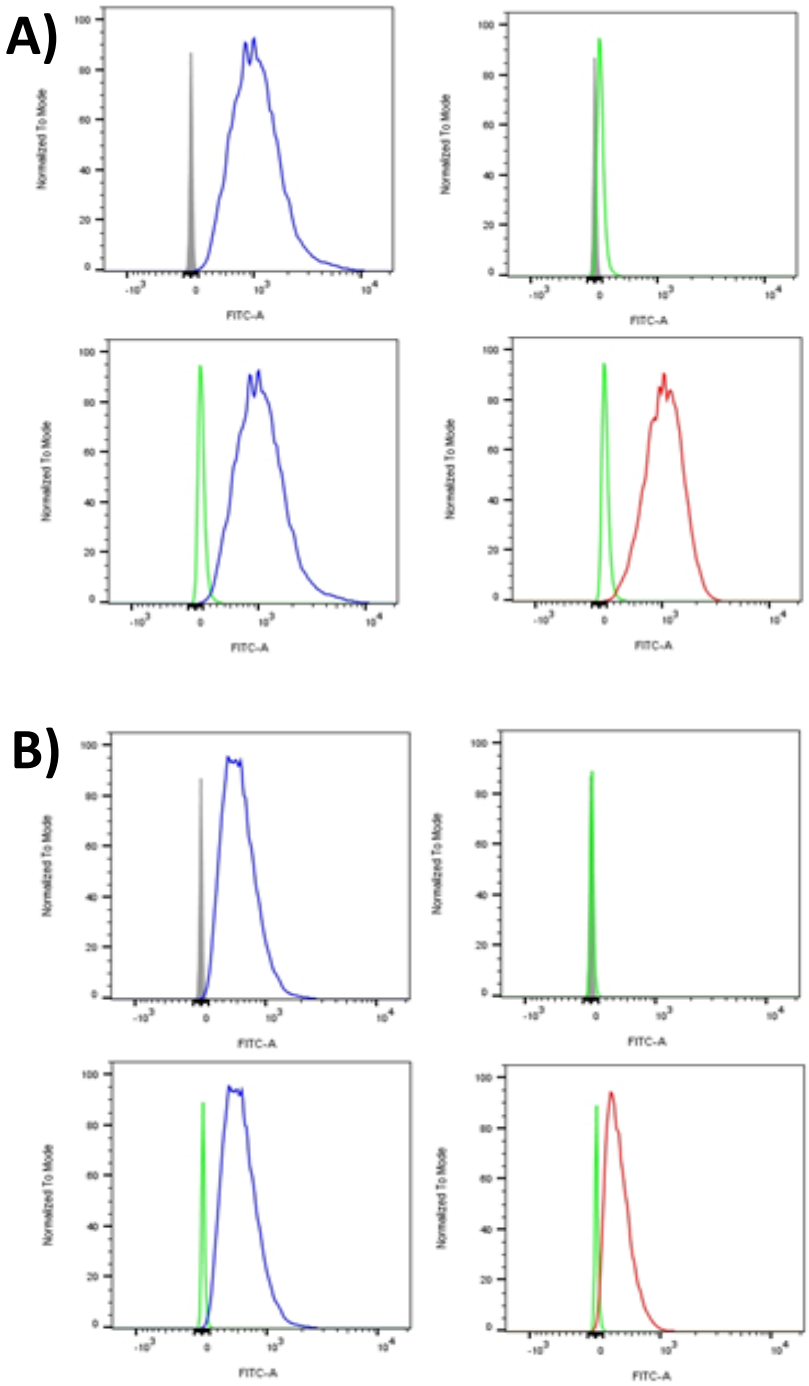
Grey = Unstained Cells

Blue = Cells probed with lectin

Green = Cells probed with lectin pre-incubated with inhibitory sugar

Red = Cells probed with lectin pre-incubated with non-inhibitory sugar

Figure 5.12: SW480 and SW620 cells probed with Jacalin and Jacalin pre-incubated with inhibitory and non-inhibitory sugars. A) SW480 cells, B) SW620 cells. Inhibitory and non-inhibitory free sugars are shown in table 5.4. Histogram colour scheme displayed in text box adjacent to A) and B).



Colour scheme for sugar inhibition histograms;

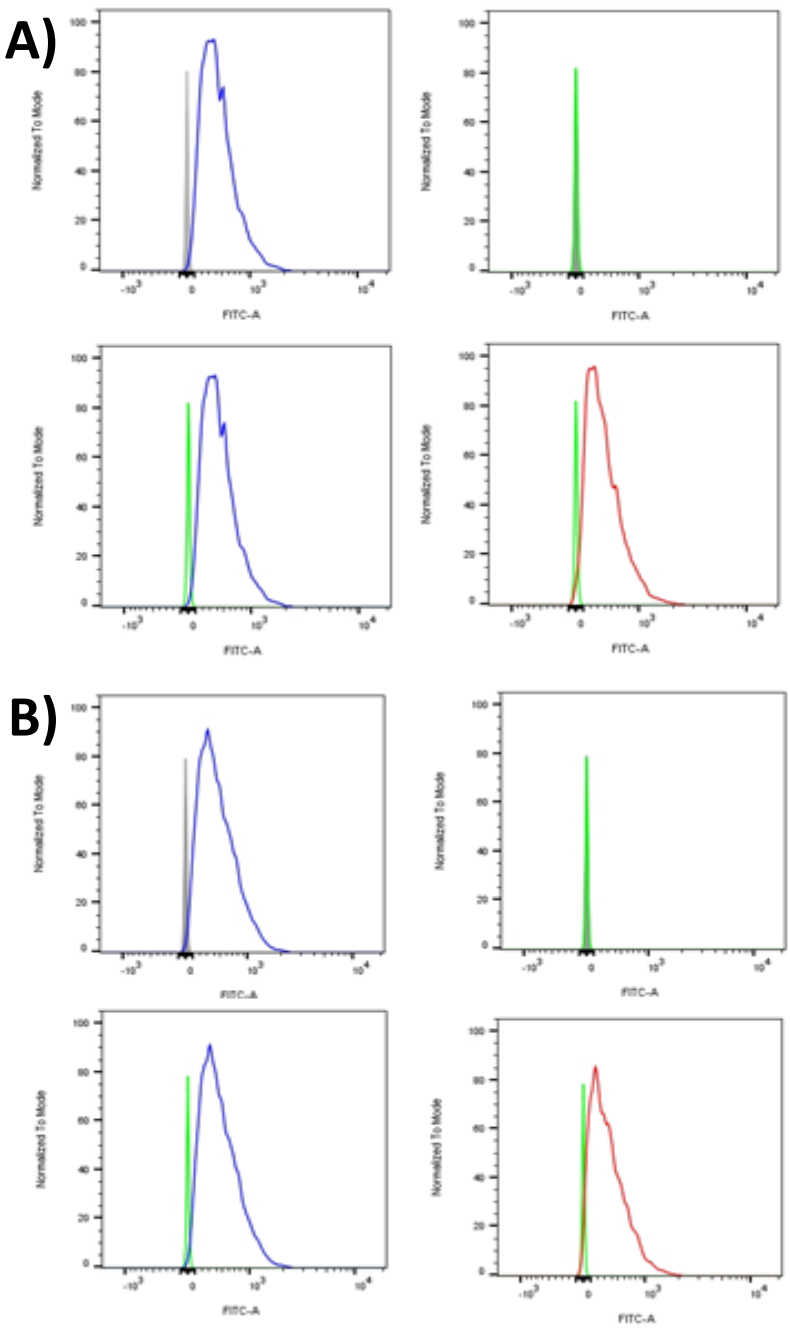
Grey = Unstained Cells

Blue = Cells probed with lectin

Green = Cells probed with lectin pre-incubated with inhibitory sugar

Red = Cells probed with lectin pre-incubated with non-inhibitory sugar

Figure 5.13: SW480 and SW620 cells probed with NPL and NPL pre-incubated with inhibitory and non-inhibitory sugars. A) SW480 cells, B) SW620 cells. Inhibitory and non-inhibitory free sugars are shown in table 5.4. Histogram colour scheme displayed in text box adjacent to A) and B).



Colour scheme for sugar inhibition histograms;

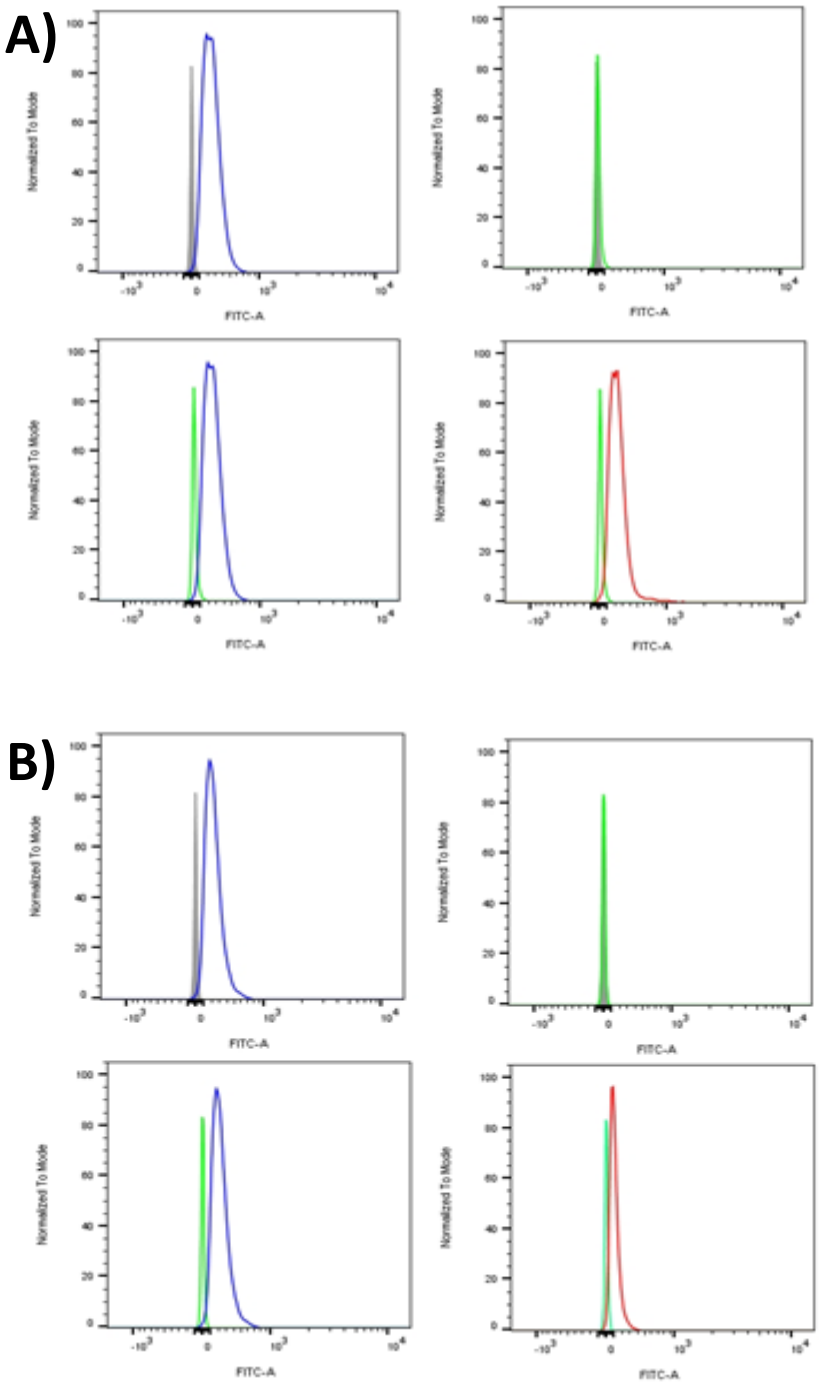
Grey = Unstained Cells

Blue = Cells probed with lectin

Green = Cells probed with lectin pre-incubated with inhibitory sugar

Red = Cells probed with lectin pre-incubated with non-inhibitory sugar

Figure 5.14: SW480 and SW620 cells probed with LecA and LecA pre-incubated with inhibitory and non-inhibitory sugars. A) SW480 cells, B) SW620 cells. Inhibitory and non-inhibitory free sugars are shown in table 5.4. Histogram colour scheme displayed in text box adjacent to A) and B).



Colour scheme for sugar inhibition histograms;

Grey = Unstained Cells

Blue = Cells probed with lectin

Green = Cells probed with lectin pre-incubated with inhibitory sugar

Red = Cells probed with lectin pre-incubated with non-inhibitory sugar

Figure 5.15: SW480 and SW620 cells probed with LecB 3K and LecB 3K pre-incubated with inhibitory and non-inhibitory sugars. A) SW480 cells, B) SW620 cells. Inhibitory and non-inhibitory free sugars are shown in table 5.4. Histogram colour scheme displayed in text box adjacent to A) and B).

5.5 Analysis of Helix Pomatia Agglutinin (HPA) binding to SW480 and SW620 cells using flow cytometry

The *Helix Pomatia* Agglutinin (HPA) lectin show some interesting results when compared to the other lectins. HPA is the only animal derived lectin used in the study and there are published reports in which it was used to distinguish between metastatic and non-metastatic cells, using both the SW480 and SW620 cell lines used in this study. Studies carried out in the past have used HPA as a negative control lectin for SW480, i.e it does not bind to these cells, and a positive marker for SW620 (Schnegelsberg et al. 2011; Saint-Guirons et al. 2007; Peiris et al. 2012). HPA is a dual specific lectin that has a strong binding affinity for both GalNAc and GlcNAc to a lesser degree (Rambaruth et al. 2012; Sanchez et al. 2006). It was understood to be solely a GalNAc binder but now is now known to bind both of these sugars and can be fully inhibited when pre-incubated with either GalNAc and GlcNAc at a concentration of 200 mM (see Figure 5.16). HPA was used to probe both SW480 and SW620 cells and the cells were viewed using fluorescent microscopy, see previous chapter section 4.3, and flow cytometry. HPA showed low level binding for SW480 cells on the fluorescent microscope, which was originally not expected, whereas it showed strong binding to some SW620 cells and no binding to others. The profile shown on fluorescent microscopy for SW620 cells fits a similar binding profile to other studies (Peiris et al. 2012). The experiment was repeated a number of times for SW480 cells and the outcome remained the same, low level binding is oberserved and HPA pre-incubated with a competitive inhibitory free sugar, GlcNAc, the binding is not observed. HPA was used to probe cells and the cells run on the flow cytometer to quantitiatively analyse the HPA binding to SW480 and SW620. Firstly HPA was tested using the ELLA platform, see section 2.24, to ensure activity and test for sugar inhibition. In figure 5.16 the HPA lectin is tested on a selection of glyco-proteins. The HPA is also pre-incubated with 200 mM of GlcNAc and 200 mM GalNAc to show binding inhibition.

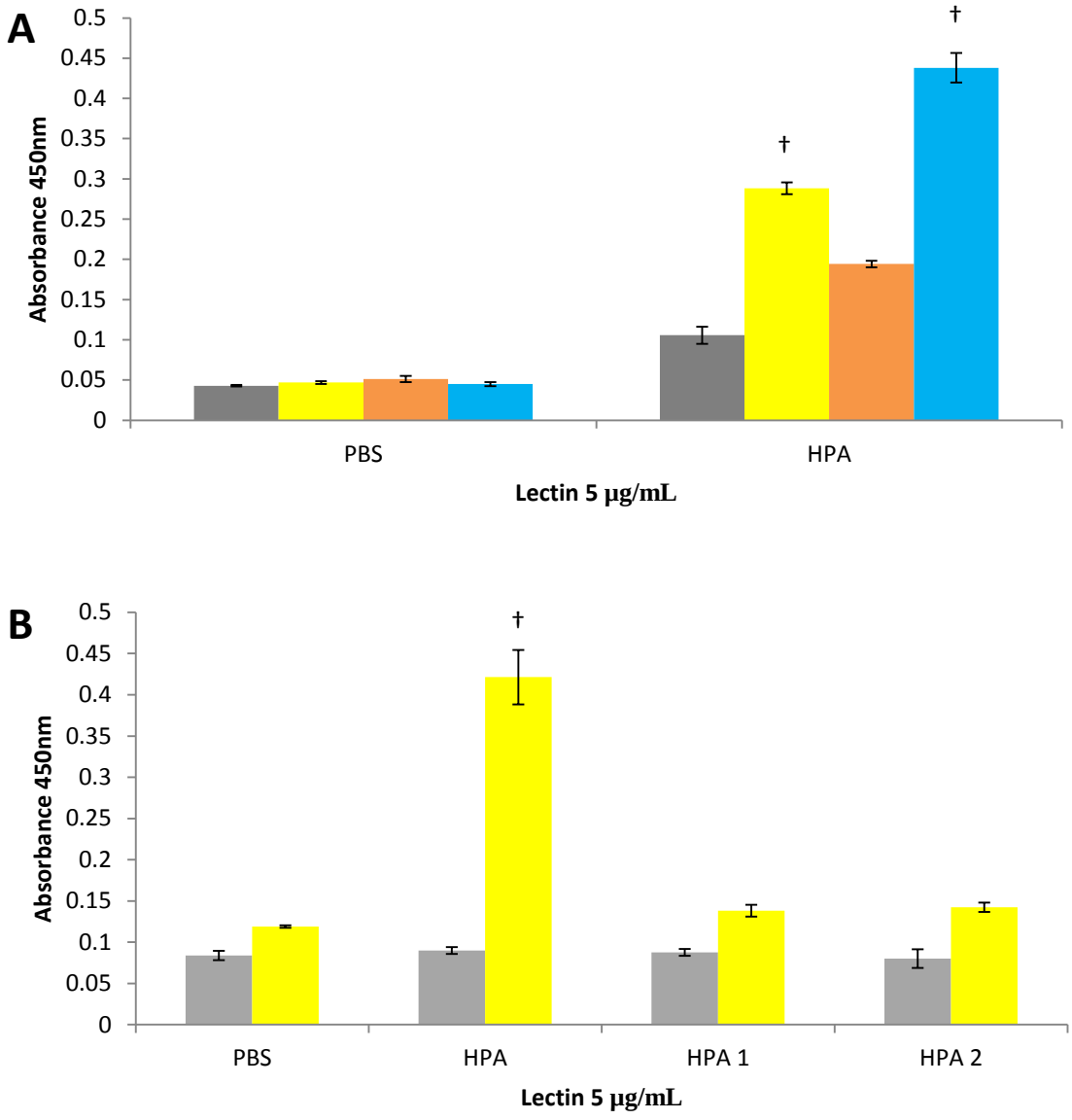


Figure 5.16: ELLA analysis of HPA binding to commercial glycoproteins including competitive sugar inhibition. A) Asialofetuin, fetuin and thyroglobulin bound with the HPA lectin, B) Asialofetuin bound with the HPA lectin as well as using competitive sugars to inhibit HPA binding. HPA 1 representes HPA pre-incubated with 200 mM GlcNAc and HPA 2 representes HPA pre-incubated with 200 mM GalNAc. Bars on the chart; Grey = PBS, Yellow = Asialofetuin, Orange = Fetuin, Light blue = Thyroglobulin. † denotes a positive value surpassing the threshold limit (Section 2.30), n = 3.

Following the same protocol and parameters laid out in Sections 2.28 and 5.2 HPA was used to probe both SW480 and SW620. Figure 5.17 shows the binding profiles of HPA bound to SW480 and SW620 analysed through flow cytometry. After testing HPA binding on both cell types, the HPA was pre-incubated with 200 mM GlcNAc to ensure binding was lectin-glycan specific. In figure 5.18 the histograms displayed show the complete sugar inhibition of HPA.

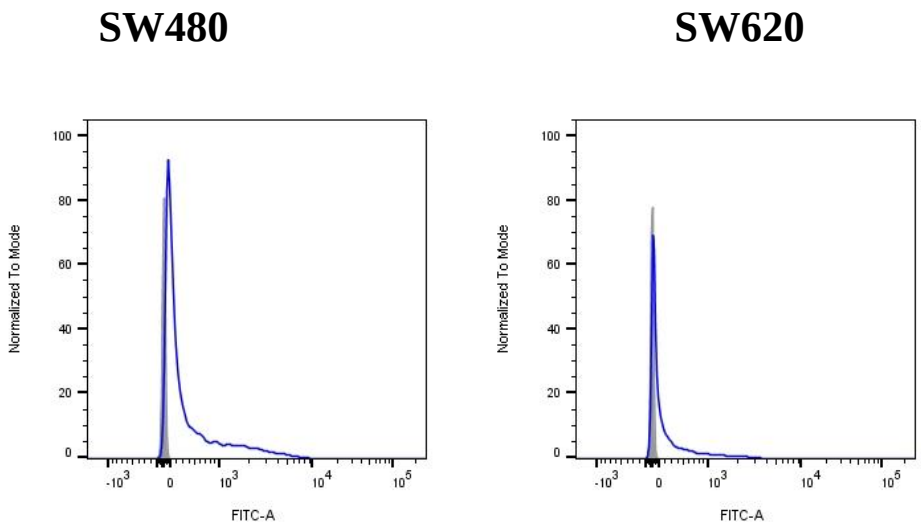


Figure 5.17: Cell surface glycoprotein analysis using HPA probe to both SW480 and SW620 cells. HPA probed cells are represented by non-filled dark blue histograms and the unstained cell population are represented by filled grey histograms. The histograms have been normalised to mode, n = 3.

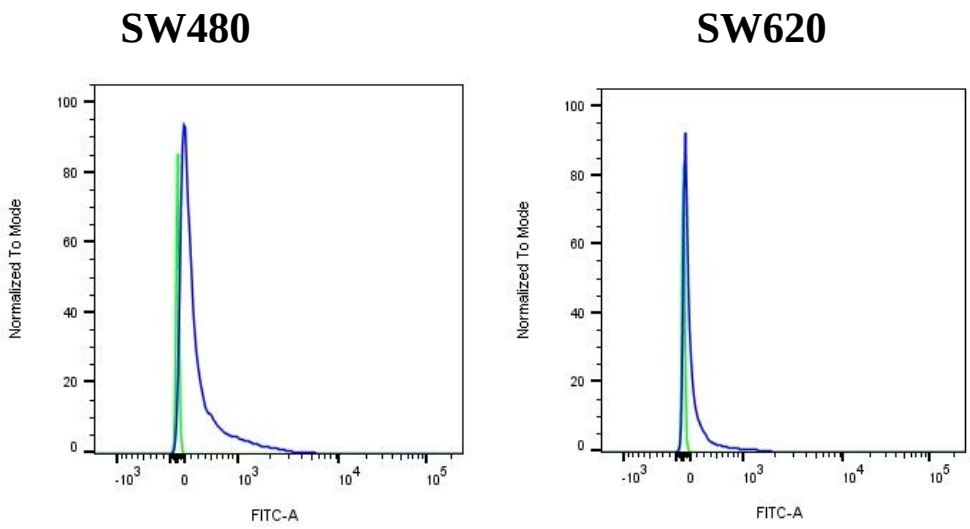


Figure 5.18: Analysis of HPA and HPA pre-incubated with 200 mM GlcNAc binding to SW480 and SW620 cells. The histograms show HPA probed cells as represented by non-filled dark blue histograms and cells probed with HPA pre-incubated with 200 mM GlcNAc are represented by non-filled green histograms. The histograms have been normalised to mode, n = 3.

Eukaryotic lectins can have batch variation due to their source. This was hypothesised as one of the possible reasons for possible HPA binding to SW480, so another batch of HPA was purchased and used to probe the SW480 and SW620 cells. A second batch of HPA showed a similar binding profile for the two cell types, see figure 5.19.

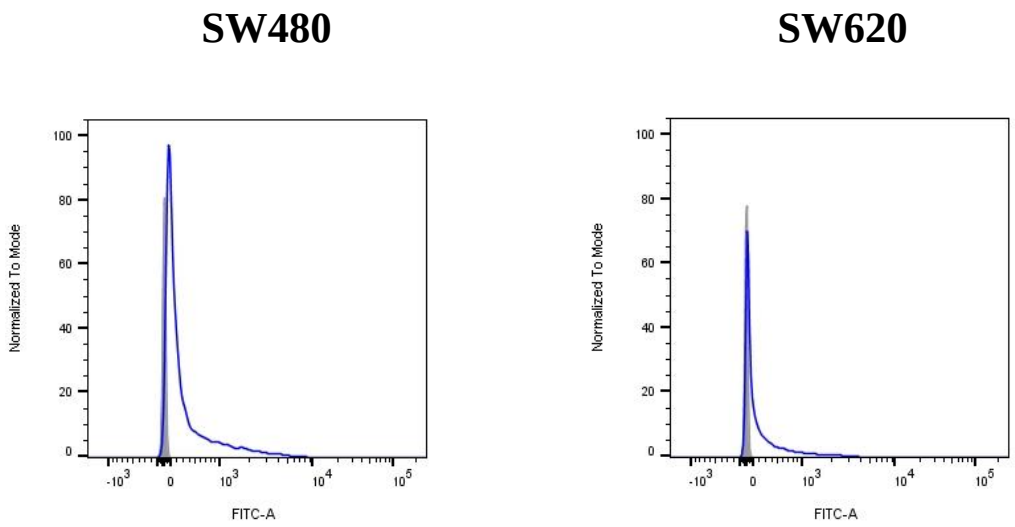


Figure 5.19: Analysis of a second batch of HPA probing to both SW480 and SW620 cells. The histograms show HPA probed cells as represented by non-filled dark blue histograms and the unstained cell population are represented by filled grey histograms. The histograms have been normalised to mode, n = 3.

A number of studies that use HPA as a negative control lectin for SW480 cells used paraformaldehyde fixed cells to probe with lectins. All probing is carried out on live cells throughout this project to ensure lectin binding was specifically on cell surface glycans only. SW480 cells were fixed to try and identify if cell fixation leads to a removal of the HPA binding observed. Both batches of HPA were used to probe fixed SW480 cells. It can be seen that the level of HPA binding increases when compared to HPA binding of non-fixed live cells, this can be seen by the reduced tailing effect in the histograms generated in figure 5.20.

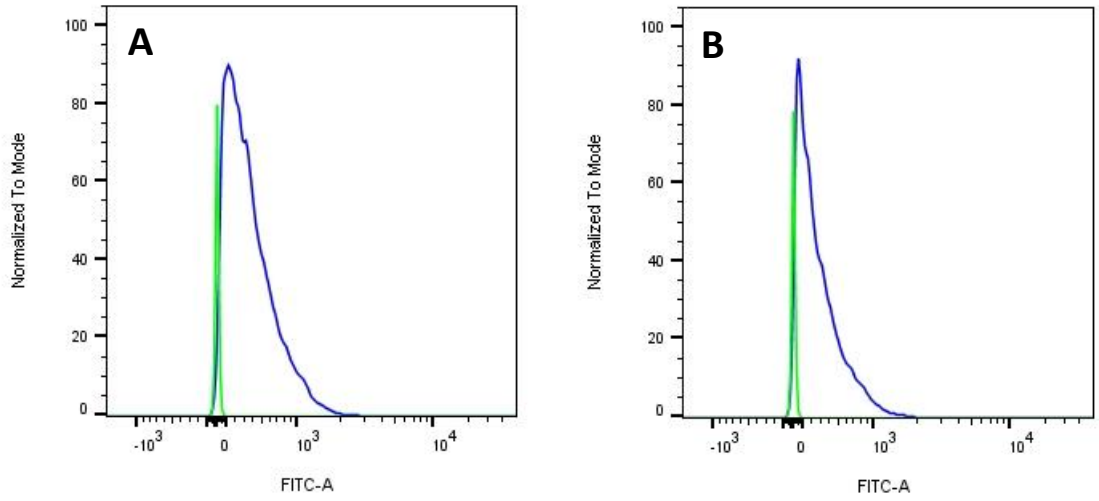


Figure 5.20: Histogram analysis of HPA binding to fixed SW480 cells. A) Original batch of HPA compared with same batch of HPA pre-incubated with 200 mM GlcNAc, B) Second batch of HPA compared with same batch of HPA pre-incubated with 200 mM GlcNAc. HPA probed cells are represented by non-filled dark blue histograms and the cells probed with HPA that had been pre-incubated with 200 mM GlcNAc are represented by non-filled green histograms. The histograms have been normalised to mode, n = 2.

After testing the difference between fixed cells and non-fixed cells the HPA lectins were still found to bind to SW480 cells. The next control experiment carried out used another batch of SW480 cells that were originally purchased from a different lab group. The cells had not been co-incubated with the SW480 cells used in this project and were never in the same cell culture room or laboratory. The cells supplied were probed with the original HPA lectin and no sugar inhibition was capable of being carried out at the time. Figure 5.21 shows the histograms obtained from the HPA binding found on these alternate SW480 cells. The binding profile is slightly different but there again is a smaller number exhibiting high level binding and a large number exhibiting low level binding.

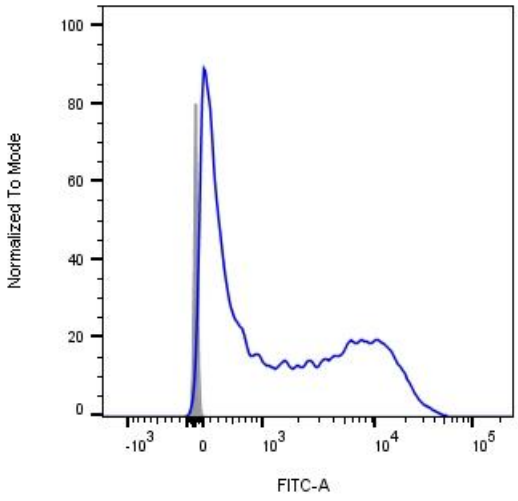


Figure 5.21: Histogram analysis of HPA binding to alternate SW480 cells. HPA probed cells are represented by non-filled dark blue histograms and the unstained cell population are represented by filled grey histograms. The histograms have been normalised to mode, n = 1.

The SW480 and SW620 cell lines used in this project were authenticated by STR genotyping (Eurofin Genomics) and routine mycoplasma testing was carried out to ensure the cells were the correct cell lines and contamination free. These controls were all carried out and the HPA probes used in this study are found to bind to SW480 repeatedly with the same binding profile exhibited in each case.

5.6 Discussion

The growth time study showed cells grown for 48 hours after passage were the optimal time for lectin probing based on the binding of a small panel of different lectins at the same concentration. The concentration study using four different lectins, two cytotoxic and two non-cytotoxic, allowed for the optimal concentration of 5 µg/mL to be worked out. The total number of cells used per sample was between 300,000-500,000 cells based on experiments carried out using the lectin Con A at 5 µg/mL. These parameters were decided based on cell viability and mean fluorescent intensity observed using both cell liness and were used for the remainder of the experiments carried out.

Flow cytometric analysis of these two cell types probed with different eukaryotic and prokaryotic lectins showed that there are significant differences in the glycans present on the cell surface. A number of lectins fit the 'normal' ratio of binding between the two cell types, see figure 5.3. Some of the lectins, both the recombinantly produced and commercially purchased lectins, showed a relative increase in certain target glycan residues on the surface of each cell type. Each cell is probed with the same concentration of lectin and the relative fluorescent intensity value of the fluorophore is recorded as well as giving the visual histograms depicting the spread of binding throughout the cell population. PNA, LCA, MAL I and ECL show a 'tailing' effect due to a larger population having low fluorescent intensities compared to some cells in the population that are exhibiting high fluorescent intensities. LCA, MAL I and ECL showed relative consistency when binding to the cells. PNA was not used in determining the ratio of binding between each cell type due to the variance displayed between each experiment. As mentioned in the previous sections PNA can bind as little as 40% and as high as 70% of cells depending on the experiment chosen. The heterogeneity in the cell lines is evident with this result.

Some lectins do not bind to either of these cell lines, such as GSL I and GSL II. There are no lectins used in this study that bind to one cell line and do not bind to the other. The lectins chosen have all been shown to bind mammalian cells. Three of lectins, GNL, UEA and MAL II, showed a significant increase in fluorescent intensity for SW480 over SW620. Their specific glycan sights are bound more on the surface of the SW480 cells indicating the levels of these glycans are increased. Five of lectins, LecA, SNA, SBA, Jacalin and MAL I, showed a significant increase in fluorescence intensity for SW620 over SW480 based on their ratio binding. SNA, SBA and Jacalin show increased binding on SW620, whereas LecA and MAL I show a near 1:1 ratio of binding between the 2 cell types. The histograms shown in figures 5.6 and 5.7 represent the binding profiles of each lectin for each cell type. The two cell types exhibit very similar histogram binding profiles. The similarities in the binding profiles give strength to the differences shown when comparing the binding ratios between the two cell lines.

A decrease in MAL II binding to SW620 is evidence of a decrease in the amount of α-2,3 linked sialic acid. This leads to an increase in the available galactose binding sites on SW620 which allows for increased LecA and Jacalin binding on SW620. SNA is also a sialic acid binder but it prefers the α-2,6 linked sugar. The increase in SNA binding coupled with the significant decrease in MAL II binding highlights a shift in the type of sialic acid capping occurring on the surface glycans of SW620 cells when compared to SW480 cells.

Along with determining the glyco-profiles of each cell line this method can also highlight small subsets of cells with different glyco-profiles. If these subsets of cells exhibit a unique glycan on their cell surface, lectins can detect these glycans using flow cytometry. DBA bound a low percentage of SW620 cells and an even lower percentage of SW480 cells, see figure 5.9, showing this lectins ability to distinguish between cells even if the percentage of positive cells is relatively low in the whole population. DBA has a binding affinity for α-linked *N*-acetylgalactosamine which is the common starting sugar for O- linked glycans but rarely found terminally α-linked.

To prove the lectins were binding their specific glycan targets a number of lectins were chosen for inhibitions studies using a corresponding inhibiting sugar as well as a non-inhibitory sugar. In section 5.4 each lectin chosen shows almost complete inhibition for each lectin chosen, with the exception of jacalin, when pre-incubated with its competitive inhibiting sugar. These lectins are also not inhibited when pre-incubated with a non-inhibitory sugar proving the specific lectin-glycan interaction on the cell surfaces of both SW480 and SW620.

HPA probing of SW480 and SW620 cells showed a similar binding pattern to the lectins PNA, MAL I, LCA and ECL with a tailing effect where some cells show low or no fluorescence and some cells show high levels of fluorescence. As explained in section 5.5 the HPA lectin was originally added to the panel of lectins used to act as a negative control for SW480 cells and show a difference in binding profiles between the two cell types. After testing different batches of HPA, SW480 cells from a different original cell stock and comparing fixed cells and non-fixed cells, all coupled with sugar inhibition studies, the binding of HPA to SW480 cells was validated. HPA bound to SW620 cells as expected and previously shown in literature (Peiris et al. 2012).

Flow cytometric analysis, through use of binding ratios as well as histogram generation, allows for more quantitative analysis of lectin binding to the surface of cells. The histograms generated show the spread of binding within a cell sample, i.e. the peak area. This along with the ratio of binding of each lectin between the two cell types allow for comprehensive analysis of different cells using a large panel of lectins where the results can be compared and contrasted to identify potential glycan changes.

Chapter 6

Final Discussion and Conclusions

In summary, the major aim of this thesis was to assess the feasibility of utilising a combination of lectin based probes and flow cytometry to interrogate the surface glycosylation of a cell. A key part of this study was to exploit the adaptability of recombinant probes. Therefore, this work initially describes the DNA manipulation carried out on the PA-IIL gene, which encodes for the LecB protein, to generate a number of altered LecB probes. This was carried out in an attempt to alter specific attributes of the protein and generate a library of probes that could be used for *in vitro* cell analysis. A number of mutations were introduced to the gene. The LecB variants described in chapter 3 were all purified and characterised to show if there were any changes in the binding specificity of the different probes. The LecB S24T mutant was an unsuccessful attempt at trying to create a LecB probe that is specific for a single sugar, which is desirable over the current dual specific wild type LecB. A knock out mutation was also added in the form of LecB S23A;S24A;G25A (LecB 3A) which was trying to alter the sugar binding pocket of the protein without impacting the calcium binding pocket which shares some of the other residues found in the sugar binding pocket. These two probes did not achieve their intended purpose. The LecB 3A protein did not reduce the binding when tested against the wild type LecB. The LecB S24T mutant also remained active and bound to both fucose and mannose with relatively similar affinity as the wild type LecB. Both of these probes were not used in the subsequent chapters 4 and 5.

The two other LecB variants were generated to use with the *in vitro* cell analysis carried out in chapters 4 and 5. Three lysine residues were added into the LecB protein between the poly-histidine tag and the start codon for the LecB gene. The other LecB variant generated was an eGFP-LecB fusion. The three lysine residues were added to aid in the biotinylated of the LecB protein which utilises the NH^{3+} side group of lysine to attach biotin labels. The eGFP-LecB protein was generated to bypass the need for direct chemical fluorophore labelling. Both of these LecB mutants proved to be potentially valuable research tools for *in vitro* cell surface analysis when used in conjunction with both fluorescent microscopy and flow cytometry.

This work has highlighted the malleability that some prokaryotic lectins show when trying to develop more unique lectin probes. Introducing a number of mutations to the wild type LecB protein allowed for the generation of a number of LecB variants. Though some of these mutations did not change the binding capabilities of the protein there are other potential mutations that could be introduced that could alter the specific binding of the lectin.

In chapters 4 and 5 the *in vitro* analysis of probing non-fixed live cells with a panel of eukaryotic and prokaryotic lectins is carried out. The panel of lectins chosen was intended to try and cover the majority of important sugars involved in mammalian glycosylation. The main sugars involved in mammalian glycosylation are explained throughout section 1.2. There are 6 pivotal sugars that are present in varying amounts between the different types of glycosylation. Fluorescent microscopy was used for qualitative analysis of lectin binding to the cell surface, as well as validation for the prokaryotic probes generated and characterised in chapter 3. After the qualitative approach was used the lectin probes were used to probe cells in suspension and the samples analysed using flow cytometry. Flow cytometry is a more quantitative method to analysis cell surface binding of lectin probes. It was the intention of this research project to show that lectin probes could be used to establish a glycan-profile for the surface glycans present on each cell line. Once the glycan-profiles were outlined, these cell types could then be compared to each other to try and identify the changes in surface glycosylation.

One problem encountered throughout chapter 5 is the potential cytotoxicity of lectins when used to probe live cells. The study carried out to try and find the optimal probing concentration shows that high levels of binding can be achieved with 5 µg/mL of lectin but even cells probed at this concentration with some lectins, such as AAL and WGA, still exhibit a high level of cell death when compared to the unstained population. The method of live, non-fixed cell probing was still chosen as the optimal way to interrogate the cell surface despite the high level of lectin toxicity. Using flow cytometry a large

number of cells can be analysed in a short period of time which can allow for statistically valid experiments to be carried out even with the potential cell lose.

The most significant problem faced is normalising for the difference in relative cell size. The cells are notably different in size when compared using the forward scatter area parameter. The difference in relative cell size is believed to equate to a difference in cell surface area due to the cells being spherical when run through a flow cytometer. Figure 5.2 shows a significant different in relative size between the two cell lines using FSC-A. The increased cell surface area of SW480 cells over SW620 cells means that a proportional increase in mean fluorescent intensity is expected. The surface of the cells is covered in glycoproteins and glycolipids that increase with the increase in surface area.

In chapter 5 the lectins that displayed a ratio of 2:1 between SW480:SW620 cells were chosen as the relative 'normal' pattern of binding that did not highlight any significant changes in these sugar residues. Three lectins showed a larger increase than 2:1 when compared between the two cell types, these were; GNL, UEA-I and MAL II. GNL binds to high and core α-mannose, UEA-I binds to α-1-2-Fucose and MAL II binds to α-2,3-NeuNAc (sialic acid). Five lectins showed a different ratio of binding between the cell lines which was closer to a 1:1 ratio. These lectins were; LecA, SNA, SBA, Jacalin and MAL I. LecA and Jacalin bind to galactose residues, with Jacalin also binding to T-antigen. SBA binds to both α- and β- linked terminal GalNAc and to a lesser extent galactose. MAL I also binds galactose but with the linkage β-1,4-GlcNAc and not β-1,3-GlcNAc which represents T-antigen. SNA binds to α-2,6-NeuNAc.

These lectins highlight changes that relate to each other. The large decrease in MAL II binding indicates a reduction in the amount of sialic acid capping occurring on the now terminal galactose in SW620 cells. This allows for an increase in galactophilic lectins to bind to the exposed galactose. An increase in SNA also indicates an increase in sialic acid capping but with an α-2,6-NeuNAc cap rather than α-2,3-NeuNAc on SW620 cells. The differential binding of these lectins highlight the cell surface glycan alterations

found on these cancer cell lines at different stages. With this, other lectins such as PNA and DBA show that subsets of cells can be identified based on their surface glycans. These findings highlight the feasibility of using flow cytometry coupled with lectins to identify quantitative differences in surface glycans between different cell lines. Ultimately using other methods to also assess the absolute cell size would allow for more sensitive quantification of lectin binding to show lower increases or decreases.

Gene expression analysis carried out on a number of different types of cancer highlight changes in glycan sialyation. In the literature a number of glycotransferase genes involved in sialyation have differential expression depending on the type of cancer tested. For instance, breast cancers have shown an increase in β-galactoside α-2,3-sialyltransferase I and VI (ST3 Gal I/VI) expression, leading to an increase in α-2,3-sialic acid capping (Pearce and Laubli 2016). This overexpression can also be seen in a number of different types of cancer such as cervical cancer (Wang 2004). Our findings showed that there was a significant decrease in α-2,3-sialylation when comparing the non-metastatic SW480 to the metastatic SW620 but also showed an increase in α-2,6-sialylation. Colorectal carcinomas have been shown in the past to have an up-regulation of in β-galactoside α-2,6-sialyltransferase (ST6 Gal I) which differs from the other two cancers mentioned. In various studies carried out ST6 Gal I is found to be upregulated while ST3 Gal genes are found to be downregulated in colorectal carcinomas (Park and Lee 2013; Dall'Olio et al. 1997). Our findings of decreased MAL II binding and increased SNA binding along with an increase in binding of a number of galactophilic lectins are in broad agreement with that as reported in the literature.

Also in chapters 4 and 5 the binding of the HPA lectin is tested against the two cell lines, with SW480 cells also being fixed to replicate some of the methods shown in literature. The binding of HPA was not expected for SW480 cells and was originally identified as a unique lectin that could distinguish between the non-metastatic SW480 cells and the metastatic SW620 cells. After testing HPA changing a number of the parameters the binding was validated and found to bind to SW480 cells both fixed and non-fixed. The control experiments included; using a second batch of HPA, using SW480 cells from

another source, using free sugars to competitively inhibit HPA binding and fixing the cells using paraformaldehyde. The two cell lines were also authenticated and the HPA binding remained constant throughout a number of repeat experiments.

All experimental results for chapters 4 and 5 generated quantitative and qualitative data for lectin binding to both cell types. Table 6.1 highlights the lectins that bound and did not bind to SW480 and SW620 cells as well as highlighting the level of binding and lectins that showed significant differences in binding between the two cell lines.

Lectin	Quantitative SW480	Quantitative SW620	Significance marker	SW480 Qualitative (+/-)	SW620 Qualitative (+/-)
AAL	++	++	*	+	+
ConA	++	++	*	+	+
DBA	+/-	+/-	N/A	+	+
DSL	++	++	-	+	+
ECL	+	+	*	+	+
GNL	++	+	*	+	+
GSL I	-	-	N/A	-	-
GSL II	-	-	N/A	-	-
HPA	+/-	+/-	N/A	+/-	+/-
Jacalin	++	++	-	+	+
LCA	+	+	*	+/-	+/-
MAL I	+	+	-	+	+
MAL II	++	+	*	+	+
NPL	++	++	*	+	+
PNA	+/-	+/-	N/A	+/-	+/-
RCA	++	++	-	+	+
SBA	++	++	-	+	+
SNA	++	++	-	+	+
UEA I	++	+	*	+	+
WGA	++	++	N/A	+	+
AAL 2	+	+	N/A	+	+
LecA	++	++	-	+	+
LecB	+	+	*	+	+

Table 6.1: Comparative results between SW480 and SW620 bound with eukaryotic and prokaryotic lectins. For quantification ++ denotes a mean fluorescent intensity value (FITC-A) of >500, + denotes a mean fluorescent intensity value (FITC-A) of 50-500 and – denotes a mean fluorescent intensity value (FITC-A) of <50. For quantification +/- denotes to lectins that produced a mixed histogram with both FITC-A positive and FITC-A negative cells. Not all lectins were compared between cell lines, and these lectins have no significance marker denoted N/A. * denotes P value <0.05 and – denotes a P value of >0.05. For qualitative results + denotes positive lectin binding, - denotes no lectin binding and +/- denotes a mix of lectin bound and unbound cells in a single population.

The relative size difference between SW480 and SW620 cells make direct comparisons more difficult, but as explained in chapter 5 there were some lectins that show differential binding between the cells. These lectin binding profiles were statistically compared to each other to show significant changes in binding between the SW480 and SW620 cells. The lectins that show increased binding to SW480 over SW620 were; GNL, UEA I and MAL II. The lectins that show increased binding to SW620 over SW480 were; LecA, SNA, SBA, Jacalin and MAL I. The lectins that show similar increased binding to SW480 over SW620, signifying no significant change in the relative amount of glycans on the surface of each cell line, were; LecB, AAL 2, AAL, Con A, DSL, ECL, RCA, LCA, NPL.

Determining the patterns of surface glycosylation on cancer cells, as well as any other cells, is a difficult process. There are a number of techniques currently used, such as lectin microarrays and mass spectrometry, which are employed to identify these surface glycans. The use of lectin probing coupled with flow cytometry is an under-utilised technique that can be used for both qualitative analysis and quantitative analysis of cell surface glycans. A large panel of lectins can be used in a single experiment with either cell lines or disaggregated tissue samples. With the continual advancements being achieved in the field of flow cytometry, this method for interrogation of glycans could be used as a reliable and useful diagnostic tool.

If there was further time and resources to develop this work there are a number of areas that would be useful to explore. There were other residues that were highlighted as potential candidates for mutagenesis in the LecB gene. The LecB variant library should be expanded, utilising of these alterations. For the *in vitro* analysis there are a number of other lectins that could be added to the panel of recombinant lectins. This expanded panel could then be utilised to identify additional changes in surface glycosylation not detected in this work. It would also be useful to expand the range of colorectal carcinoma cell lines in order to try and identify if there are other significant changes in lectin binding patterns between cells from different patients. It would also be useful to use some more powerful analytical instrumentation. For example, the use of a state of

the art confocal microscope or imaging flow cytometer that would allow for more in depth analysis of lectin binding to the cell surface to be carried out.

In conclusion, this research project allowed for the mutagenesis, expression, purification and characterisation of the prokaryotic lectin LecB. Some of these probes were engineered to perform better when used *in vitro* on SW480 and SW620 cells. Once characterised the probes were validated for *in vitro* use and then used for probing cells and analysing binding through fluorescent microscopy and flow cytometry. Fluoroescent microscopy and flow cytometry were both used for qualitative analysis of cell surface glycosylation and flow cytometry could also be used for quantitative analysis. The glycan pattern on the surface of both SW480 and SW620 cells were assessed and then compared to show the difference in surface glycans. The original aims at the beginning of the project were achieved.

Chapter 7

References

Abcam. *Introduction to flow cytometry* [Online]. Available from: http://docs.abcam.com/pdf/protocols/flow-cytometry-introduction.pdf [Accessed 2/20 2017].

Abhilash, J., Dileep, K.V., Palanimuthu, M., Geethanandan, K., Sadasivan, C. and Haridas, M. 2013. Metal ions in sugar binding, sugar specificity and structural stability of spatholobus parviflorus seed lectin. *Journal of Molecular Modeling*, 19(8), pp.3271-3278.

Aebi, M. 2013. N-linked protein glycosylation in the ER. *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research*, 1833(11), pp.2430-2437.

Ahmad, S., Khan, R.H. and Ahmad, A. 1999. Physicochemical characterization of cajanus cajan lectin: Effect of pH and metal ions on lectin carbohydrate interaction. *Biochimica Et Biophysica Acta (BBA)-General Subjects*, 1427(3), pp.378-384.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. 2002. Molecular biology of the cell, 4th edition. New York: Garland Science

American Type Culture Collection. *Thawing, propagating, and cryopreserving protocol* [Online]. Available from: https://physics.cancer.gov/docs/bioresource/colorectal/NCI-PBCF-CCL228_SW480_SOP-508.pdf [Accessed 05/05 2017].

Basakran, N.S. 2015. CD44 as a potential diagnostic tumor marker. *Saudi Medical Journal*, 36(3), pp.273-279.

Batisse, C., Marquet, J., Greffard, A., Fleury-Feith, J., Jaurand, M. and Pilatte, Y. 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.

Bennett, P. 2008. Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. *British Journal of Pharmacology,* 153(S1), pp.S347-S357.

Bieberich, E. 2014. Synthesis, processing, and function of N-glycans in N-glycoproteins. *Glycobiology of the Nervous System.* Springer, 9, pp.47-70.

Biosciences, B. D. 2000. Introduction to flow cytometry: A learning guide. *Manual Part,* 1.

Boudabbous, M., Hmad, I.B., Saibi, W., Mssawra, M., Belghith, H. and Gargouri, A. 2016. Trans-glycosylation capacity of a highly glycosylated multi-specific β-glucosidase from fusarium solani. *Bioprocess and Biosystems Engineering,* 40(4), pp.1-13.

Bovin, N. 2013. Natural antibodies to glycans. *Biochemistry (Moscow),* 78(7), pp.786-797.

Brenstrum, T.J. and Brimble, M.A. 2001. C-glycosylation of naphthols using glucosyl donors. *Arkivoc,* 7, pp.37-48.

Brooks, S.A. 2000. The involvement of helix pomatia lectin (HPA) binding N-acetylgalactosamine glycans in cancer progression. *Histology and Histopathology,* 15(1), pp.143-158.

Brooks, S.A. and Carter, T.M. 2001. N-acetylgalactosamine, N-acetylglucosamine and sialic acid expression in primary breast cancers. *Acta Histochemica,* 103(1), pp.37-51.

Campbell, M.P., Royle, L., Radcliffe, C.M., Dwek, R.A. and Rudd, P.M. 2008. GlycoBase and autoGU: Tools for HPLC-based glycan analysis. *Bioinformatics (Oxford, England),* 24(9), pp.1214-1216.

Cancer Research, UK [Online]. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/survival#heading-Three [Accessed 4/24 2017].

Cao, J., Guo, S., Arai, K., Lo, E.H. and Ning, M. 2013. Studying extracellular signaling utilizing a glycoproteomic approach: Lectin blot surveys, a first and important step. *Chemokines: Methods and Protocols,* 1013, pp.227-233.

Chemani, C., Imberty, A., de Bentzmann, S., Pierre, M., Wimmerova, M., Guery, B.P. and Faure, K. 2009. Role of LecA and LecB lectins in pseudomonas aeruginosa-induced lung injury and effect of carbohydrate ligands. *Infection and Immunity,* 77(5), pp.2065-2075.

Christiansen, M.N., Chik, J., Lee, L., Anugraham, M., Abrahams, J.L. and Packer, N.H. 2014. Cell surface protein glycosylation in cancer. *Proteomics,* 14(4-5), pp.525-546.

Colley, K.J. and Baenziger, J.U. 1987. Identification of the post-translational modifications of the core-specific lectin. the core-specific lectin contains hydroxyproline, hydroxylysine, and glucosylgalactosylhydroxylysine residues. *The Journal of Biological Chemistry,* 262(21), pp.10290-10295.

Cui, H., Lin, Y., Yue, L., Zhao, X. and Liu, J. 2011. Differential expression of the α2, 3-sialic acid residues in breast cancer is associated with metastatic potential. *Oncology Reports,* 25(5), pp.1365-1371.

Dall'Olio, F., Mariani, E., Tarozzi, A., Meneghetti, A., Chiricolo, M., Lau, J.T. and Facchini, A. 1997. Expression of beta-galactoside alpha 2,6-sialyltransferase does not alter the susceptibility of human colon cancer cells to NK-mediated cell lysis. *Glycobiology,* 7(4), pp.507-513.

de Freitas Junior, J.C.M., Bárbara Du Rocher, D., de Souza, W.F., de Araújo, W.M., Abdelhay, E.S.F.W. and Morgado-Díaz, J.A., 2011. Inhibition of N-linked glycosylation by tunicamycin induces E-cadherin-mediated cell–cell adhesion and inhibits cell proliferation in undifferentiated human colon cancer cells. *Cancer chemotherapy and pharmacology*, *68*(1), pp.227-238.

de Freitas Junior, J.C. and Morgado-Diaz, J.A. 2016. The role of N-glycans in colorectal cancer progression: Potential biomarkers and therapeutic applications. *Oncotarget*, 7(15), pp.19395-19413.

Dommett, R., Klein, N. and Turner, M. 2006. Mannose-binding lectin in innate immunity: Past, present and future. *Tissue Antigens*, 68(3), pp.193-209.

Drake, R.R., Jones, E.E., Powers, T.W. and Nyalwidhe, J.O. 2015. Chapter ten-altered glycosylation in prostate cancer. *Advances in Cancer Research*, 126pp.345-382.

Dwek, R.A. 1998. Biological importance of glycosylation. Biological importance of glycosylation. *Molecular Recognition and Inclusion: Proceedings of the Ninth International Symposium on Molecular Recognition and Inclusion, Held at Lyon, 7-12 September 1996.* Springer, 96, pp. 43-47.

Franz, S., Frey, B., Sheriff, A., Gaipl, U.S., Beer, A., Voll, R.E., Kalden, J.R. and Herrmann, M. 2006. Lectins detect changes of the glycosylation status of plasma membrane constituents during late apoptosis. *Cytometry Part A*, 69(4), pp.230-239.

Fujihashi, M., Peapus, D.H., Kamiya, N., Nagata, Y. and Miki, K. 2003. Crystal structure of fucose-specific lectin from *Aleuria Aurantia* binding ligands at three of its five sugar recognition sites. *Biochemistry*, 42(38), pp.11093-11099.

Funken, H., Bartels, K., Wilhelm, S., Brocker, M., Bott, M., Bains, M., Hancock, R.E., Rosenau, F. and Jaeger, K. 2012. Specific association of lectin LecB with the surface of *Pseudomonas Aeruginosa*: Role of outer membrane protein OprF. *PloS One,* 7(10), pp.e46857.

Galdeen, S.A. and North, A.J. 2011. Live cell fluorescence microscopy techniques. *Cell Migration: Developmental Methods and Protocols*, 769, pp.205-222.

Gazit, Y., Mory, A., Etzioni, A., Frydman, M., Scheuerman, O., Gershoni-Baruch, R. and Garty, B. 2010. Leukocyte adhesion deficiency type II: Long-term follow-up and review of the literature. *Journal of Clinical Immunology,* 30(2), pp.308-313.

Gohlke, M. and Blanchard, V. 2008. Separation of N-glycans by HPLC. *Post-Translational Modifications of Proteins: Tools for Functional Proteomics,* 446, pp.239-254.

Goodison, S., Urquidi, V. and Tarin, D. 1999. CD44 cell adhesion molecules. *Molecular Pathology : MP*, 52(4), pp.189-196.

Gorakshakar, A.C. and Ghosh, K. 2016. Use of lectins in immunohematology. *Asian Journal of Transfusion Science,* 10(1), pp.12.

Grünwald-Gruber, C., Thader, A., Maresch, D., Dalik, T. and Altmann, F. 2017. Determination of true ratios of different N-glycan structures in electrospray ionization mass spectrometry. *Analytical and Bioanalytical Chemistry,* 409(10), pp.2519-2530.

Gu, J., Isaji, T., Xu, Q., Kariya, Y., Gu, W., Fukuda, T. and Du, Y. 2012. Potential roles of N-glycosylation in cell adhesion. *Glycoconjugate Journal*, 29(8-9), pp.599-607.

Hagerbaumer, P., Vieth, M., Anders, M. and Schumacher, U. 2015. Lectin histochemistry shows WGA, PHA-L and HPA binding increases during progression of human colorectal cancer. *Anticancer Research,* 35(10), pp.5333-5339.

Han, J.W., Jung, M.G., Shim, E.Y., Shim, J.B., Kim, Y.M. and Kim, G.H. 2015. Functional recombinants designed from a Fetuin/Asialofetuin-specific marine algal lectin, rhodobindin. *Marine Drugs,* 13(4), pp.2183-2195.

Hanahan, D. and Weinberg, R.A. 2011. Hallmarks of cancer: The next generation. *Cell,* 144(5), pp.646-674.

Hanahan, D. and Weinberg, R.A. 2000. The hallmarks of cancer. *Cell,* 100(1), pp.57-70.

Häuselmann, I. and Borsig, L. 2014. Altered tumor-cell glycosylation promotes metastasis. *Frontiers in Oncology*, 4, pp.28.

Herzenberg, L.A., Tung, J., Moore, W.A., Herzenberg, L.A. and Parks, D.R. 2006. Interpreting flow cytometry data: A guide for the perplexed. *Nature Immunology,* 7(7), pp.681.

Hoff, F. *How to prepare your specimen for immunofluorescence microscopy* [Online]. Available from: http://www.leica-microsystems.com/science-lab/how-to-prepare-your-specimen-for-immunofluorescence-microscopy/ [Accessed 5/12 2017].

Holmes, K., Lantz, L.M., Fowlkes, B., Schmid, I. and Giorgi, J.V. 2001. Preparation of cells and reagents for flow cytometry. *Current Protocols in Immunology*, 44, pp. 5.3:5.3.1–5.3.24.

Hsu, K., 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.

Hu, S. and Wong, D.T. 2009. Lectin microarray. *Proteomics-Clinical Applications,* 3(2), pp.148-154.

Huang, Y., Lin, Y., Shih, C., Shih, C., Chang, T. and Chen, C. 2006. Purification, crystallization and preliminary X-ray crystallographic analysis of rice lectin from oryza sativa. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications,* 62(2), pp.94-96.

Irish Cancer Society [Online]. Available from: https://www.cancer.ie/about-us/media-centre/cancer-statistics/most-common-cancers#sthash.tvxEZKT6.dpbs [Accessed 4/3 2017].

Ise, H., Goto, M., Komura, K. and Akaike, T. 2012. Engulfment and clearance of apoptotic cells based on a GlcNAc-binding lectin-like property of surface vimentin. *Glycobiology,* 22(6), pp.788-805.

Johnson, M.B. and Criss, A.K. 2013. Fluorescence microscopy methods for determining the viability of bacteria in association with mammalian cells. *JoVE (Journal of Visualized Experiments),* 79, pp.e50729-e50729.

Julien, S., Ivetic, A., Grigoriadis, A., QiZe, D., Burford, B., Sproviero, D., Picco, G., Gillett, C., Papp, S.L., Schaffer, L., Tutt, A., Taylor-Papadimitriou, J., Pinder, S.E. and Burchell, J.M. 2011. Selectin ligand sialyl-lewis x antigen drives metastasis of hormone-dependent breast cancers. *Cancer Research,* 71(24), pp.7683-7693.

Kekalainen, J., Larma, I., Linden, M. and Evans, J.P. 2015. Lectin staining and flow cytometry reveals female-induced sperm acrosome reaction and surface carbohydrate reorganization. *Scientific Reports,* 5pp.15321.

Keogh, D., Thompson, R., Larragy, R., McMahon, K., O'Connell, M., O'Connor, B. and Clarke, P. 2014. Generating novel recombinant prokaryotic lectins with altered carbohydrate binding properties through mutagenesis of the PA-IL protein from *Pseudomonas Aeruginosa*. *Biochimica Et Biophysica Acta (BBA)-General Subjects,* 1840(6), pp.2091-2104.

Khan Academy [Online]. Available from: https://www.khanacademy.org/science/biology/membranes-and-transport/the-plasma-membrane/a/structure-of-the-plasma-membrane [Accessed 2/15 2017].

Khan, F., Khan, R.H., Sherwani, A., Mohmood, S. and Azfer, M.A. 2002. Lectins as markers for blood grouping. *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research,* 8(12), pp.RA293-300.

Kim, J.J. and Park, K. 2001. Glucose-binding property of pegylated concanavalin A. *Pharmaceutical Research,* 18(6), pp.794-799.

Kirkeby, S., Hansen, A.K., d'Apice, A. and Moe, D. 2006. The galactophilic lectin (PA-IL, gene LecA) from *Pseudomonas Aeruginosa*. its binding requirements and the localization of lectin receptors in various mouse tissues. *Microbial Pathogenesis,* 40(5), pp.191-197.

Knijnenburg, T.A., Roda, O., Wan, Y., Nolan, G.P., Aitchison, J.D. and Shmulevich, I. 2011. A regression model approach to enable cell morphology correction in high-throughput flow cytometry. *Molecular Systems Biology,* 7, pp.531.

Kobata, A. and Amano, J. 2005. Altered glycosylation of proteins produced by malignant cells, and application for the diagnosis and immunotherapy of tumours. *Immunology and Cell Biology,* 83(4), pp.429-439.

Kölbl, A.C., Andergassen, U. and Jeschke, U. 2015. The role of glycosylation in breast cancer metastasis and cancer control. *Frontiers in Oncology,* 5pp.219.

Kravitz, M.S. and Shoenfeld, Y. 2006. Autoimmunity to protective molecules: Is it the perpetuum mobile (vicious cycle) of autoimmune rheumatic diseases? *Nature Clinical Practice Rheumatology,* 2(9), pp.481-490.

Lairson, L., Henrissat, B., Davies, G. and Withers, S. 2008. Glycosyltransferases: Structures, functions, and mechanisms. *Annual Review of Biochemistry.* 77, pp.521-555.

Lam, S.K. and Ng, T.B. 2011. Lectins: Production and practical applications. *Applied Microbiology and Biotechnology*, 89(1), pp.45-55.

Leibovitz, A., Stinson, J.C., McCombs, W.B.,3rd, McCoy, C.E., Mazur, K.C. and Mabry, N.D. 1976. Classification of human colorectal adenocarcinoma cell lines. *Cancer Research,* 36(12), pp.4562-4569.

Lira-Navarrete, E., Valero-González, J., Villanueva, R., Martínez-Júlvez, M., Tejero, T., Merino, P., Panjikar, S. and Hurtado-Guerrero, R. 2011. Structural insights into the mechanism of protein O-fucosylation. *PLoS One,* 6(9), pp.e25365.

Liu, L., Telford, J.E., Knezevic, A. and Rudd, P.M. 2010. High-throughput glycoanalytical technology for systems glycobiology. *Biochemical Society Transactions,* 38(5), pp.1374-1377.

Llop, E., Ferrer-Batallé, M., Barrabés, S., Guerrero, P.E., Ramírez, M., Saldova, R., Rudd, P.M., Aleixandre, R.N., Comet, J. and de Llorens, R. 2016. Improvement of prostate cancer diagnosis by detecting PSA glycosylation-specific changes. *Theranostics,* 6(8), pp.1190-1204.

Lodish, H. and Zipursky, S.L. 2001. Molecular cell biology. *Biochemistry and Molecular Biology Education,* 29pp.126-133.

Lommel, M. and Strahl, S. 2009. Protein O-mannosylation: Conserved from bacteria to humans*. *Glycobiology,* 19(8), pp.816-828.

Loris, R., Tielker, D., Jaeger, K. and Wyns, L. 2003. Structural basis of carbohydrate recognition by the lectin LecB from *Pseudomonas Aeruginosa*. *Journal of Molecular Biology,* 331(4), pp.861-870.

Loureiro, L.R., Carrascal, M.A., Barbas, A., Ramalho, J.S., Novo, C., Delannoy, P. and Videira, P.A. 2015. Challenges in antibody development against tn and sialyl-tn antigens. *Biomolecules,* 5(3), pp.1783-1809.

Lyons, J.J., Milner, J.D. and Rosenzweig, S.D. 2015. Glycans instructing immunity: The emerging role of altered glycosylation in clinical immunology. *Frontiers in Pediatrics,* 3, pp.54.

Mahmood, T. and Yang, P.C. 2012. Western blot: Technique, theory, and trouble shooting. *North American Journal of Medical Sciences,* 4(9), pp.429-434.

Masters, J.R. 2002. HeLa cells 50 years on: The good, the bad and the ugly. *Nature Reviews Cancer,* 2(4), pp.315-319.

Maupin, K.A., Liden, D. and Haab, B.B. 2012. The fine specificity of mannose-binding and galactose-binding lectins revealed using outlier motif analysis of glycan array data. *Glycobiology,* 22(1), pp.160-169.

Maverakis, E., Kim, K., Shimoda, M., Gershwin, M.E., Patel, F., Wilken, R., Raychaudhuri, S., Ruhaak, L.R. and Lebrilla, C.B. 2015. Glycans in the immune system and the altered glycan theory of autoimmunity: A critical review. *Journal of Autoimmunity,* 57pp.1-13.

Morelle, W. and Michalski, J. 2007. Analysis of protein glycosylation by mass spectrometry. *Nature Protocols,* 2(7), pp.1585-1602.

Morrison, K.L. and Weiss, G.A. 2001. Combinatorial alanine-scanning. *Current Opinion in Chemical Biology,* 5(3), pp.302-307.

Nakayama, Y., Nakamura, N., Tsuji, D., Itoh, K. and Kurosaka, A. 2013. Genetic diseases associated with protein glycosylation disorders in mammals. Genetic Disorders, Prof. Maria Puiu editior. InTech.

National Cancer Insitute. *What is cancer?*[Online]. Available from: https://www.cancer.gov/about-cancer/understanding/what-is-cancer [Accessed 4/16 2017].

O'Connor, B.F., Monaghan, D. and Cawley, J. 2017. Lectin affinity chromatography (LAC). *Protein Chromatography: Methods and Protocols*, pp.411-420.

Ohtsubo, K. and Marth, J.D. 2006. Glycosylation in cellular mechanisms of health and disease. *Cell*, 126(5), pp.855-867.

Okajima, T. and Irvine, K.D. 2002. Regulation of notch signaling by o-linked fucose. *Cell*, 111(6), pp.893-904.

Oliveira, C., Teixeira, J.A. and Domingues, L. 2014. Recombinant production of plant lectins in microbial systems for biomedical application–the frutalin case study. *Frontiers in Plant Science*, 5, pp.390.

Ormerod, M.G. and Novo, D. 2008. *Flow cytometry: a basic introduction.* [Online]. Available from: http://flowbook.denovosoftware.com/ [Accessed 04/04 2017]

Park, J. and Lee, M. 2013. Increasing the α 2, 6 sialylation of glycoproteins may contribute to metastatic spread and therapeutic resistance in colorectal cancer. *Gut Liver*, 7(6), pp.629-641.

Paulick, M.G. and Bertozzi, C.R. 2008. The glycosylphosphatidylinositol anchor: A complex membrane-anchoring structure for proteins. *Biochemistry*, 47(27), pp.6991-7000.

Pearce, O.M. and Läubli, H. 2016. Sialic acids in cancer biology and immunity. *Glycobiology*, 26(2), pp.111-128.

Peiris, D., Markiv, A., Curley, G.P. and Dwek, M.V. 2012. A novel approach to determining the affinity of protein–carbohydrate interactions employing adherent cancer cells grown on a biosensor surface. *Biosensors and Bioelectronics,* 35(1), pp.160-166.

Peracaula, R., Tabares, G., Royle, L., Harvey, D.J., Dwek, R.A., Rudd, P.M. and de Llorens, R. 2003. Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology,* 13(6), pp.457-470.

Peterson, L.W. and Artis, D. 2014. Intestinal epithelial cells: Regulators of barrier function and immune homeostasis. *Nature Reviews Immunology,* 14(3), pp.141-153.

Pilobello, K.T. and Mahal, L.K. 2007. Lectin microarrays for glycoprotein analysis. *Microchip-Based Assay Systems: Methods and Applications,*385 pp.193-203.

Pinho, S.S., Carvalho, S., Marcos-Pinto, R., Magalhães, A., Oliveira, C., Gu, J., Dinis-Ribeiro, M., Carneiro, F., Seruca, R. and Reis, C.A. 2013. Gastric cancer: Adding glycosylation to the equation. *Trends in Molecular Medicine,* 19(11), pp.664-676.

Pinho, S.S., Reis, C.A., Paredes, J., Magalhaes, A.M., Ferreira, A.C., Figueiredo, J., Xiaogang, W., Carneiro, F., Gartner, F. and Seruca, R. 2009. The role of N-acetylglucosaminyltransferase III and V in the post-transcriptional modifications of E-cadherin. *Human Molecular Genetics,* 18(14), pp.2599-2608.

Ponta, H., Sherman, L. and Herrlich, P.A. 2003. CD44: From adhesion molecules to signalling regulators. *Nature Reviews Molecular Cell Biology,* 4(1), pp.33-45.

Praissman, J.L. and Wells, L. 2014. Mammalian O-mannosylation pathway: Glycan structures, enzymes, and protein substrates. *Biochemistry,* 53(19), pp.3066-3078.

Rambaruth, N.D., Greenwell, P. and Dwek, M.V. 2012. The lectin helix pomatia agglutinin recognizes O-GlcNAc containing glycoproteins in human breast cancer. *Glycobiology,* 22(6), pp.839-848.

Rambaruth, N.D., Greenwell, P. and Dwek, M.V. 2012. The lectin helix pomatia agglutinin recognizes O-GlcNAc containing glycoproteins in human breast cancer. *Glycobiology,* 22(6), pp.839-848.

Reeke Jr, G.N., Becker, J.W., Cunningham, B.A., Wang, J.L., Yahara, I. and Edelman, G.M. 1975. Structure and function of concanavalin A. *Advances in Experimental Medicien and Biology*, 55, pp.13-33.

Rosano, G.L. and Ceccarelli, E.A. 2014. Recombinant protein expression in escherichia coli: Advances and challenges. *Recombinant Protein Expression in Microbial Systems,* 7

Sabin, C., Mitchell, E.P., Pokorná, M., Gautier, C., Utille, J., Wimmerová, M. and Imberty, A. 2006. Binding of different monosaccharides by lectin PA-IIL from pseudomonas aeruginosa: Thermodynamics data correlated with X-ray structures. *FEBS Letters*, 580(3), pp.982-987.

Saint-Guirons, J., Zeqiraj, E., Schumacher, U., Greenwell, P. and Dwek, M. 2007. Proteome analysis of metastatic colorectal cancer cells recognized by the lectin helix pomatia agglutinin (HPA). *Proteomics,* 7(22), pp.4082-4089.

Sanchez, J.F., Lescar, J., Chazalet, V., Audfray, A., Gagnon, J., Alvarez, R., Breton, C., Imberty, A. and Mitchell, E.P. 2006. Biochemical and structural analysis of helix pomatia agglutinin. A hexameric lectin with a novel fold. *The Journal of Biological Chemistry,* 281(29), pp.20171-20180.

Sato, T. 2014. Lectin-probed western blot analysis. *Lectins: Methods and Protocols,*1200, pp.93-100.

Schaeffer-Reiss, C. 2008. A brief summary of the different types of mass spectrometers used in proteomics. *Functional Proteomics: Methods and Protocols,* 484, pp.3-16.

Schnegelsberg, B., Schumacher, U. and Valentiner, U. 2011. Lectin histochemistry of metastasizing and non-metastasizing breast and colon cancer cells. *Anticancer Research,* 31(5), pp.1589-1597.

Schumacher, U. and Adam, E. 1997. Lectin histochemical HPA-binding pattern of human breast and colon cancers is associated with metastases formation in severe combined immunodeficient mice. *The Histochemical Journal*, 29(9), pp.677-684.

Sezonov, G., Joseleau-Petit, D. and D'Ari, R. 2007. *Escherichia coli* physiology in luria-bertani broth. *Journal of Bacteriology,* 189(23), pp.8746-8749.

Shental-Bechor, D. and Levy, Y. 2008. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences of the United States of America,* 105(24), pp.8256-8261.

Simple Cloning Lab. *Electrophoresis* [Online]. Available from: http://www.bioinformatics.nl/molbi/SimpleCloningLab/electrophoresis.htm [Accessed 20/03/2017 2017].

Sirard, T. *Fundamentals of HPLC* [Online]. Available from: http://www.waters.com/webassets/cms/events/docs/FundamentalsofHPLCWebinar_TRS_102012.pdf [Accessed February 13 2017].

Spector, D.L. and Goldman, R.D. 2006. Constructing and expressing GFP fusion proteins. *CSH Protocols,* 2006(7), pp.10.1101/pdb.prot4649.

Spiro, R.G. 2002. Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology,* 12(4), pp.43R-56R.

Stalnaker, S.H., Aoki, K., Lim, J.M., Porterfield, M., Liu, M., Satz, J.S., Buskirk, S., Xiong, Y., Zhang, P., Campbell, K.P., Hu, H., Live, D., Tiemeyer, M. and Wells, L. 2011. Glycomic analyses of mouse models of congenital muscular dystrophy. *The Journal of Biological Chemistry,* 286(24), pp.21180-21190.

Stenman, U., Alfthan, H. and Hotakainen, K. 2004. Human chorionic gonadotropin in cancer. *Clinical Biochemistry,* 37(7), pp.549-561.

Su, D., Zhao, H. and Xia, H. 2010. Glycosylation-modified erythropoietin with improved half-life and biological activity. *International Journal of Hematology,* 91(2), pp.238-244.

Sukhdeo, K., Paramban, R.I., Vidal, J.G., Elia, J., Martin, J., Rivera, M., Carrasco, D.R., Jarrar, A., Kalady, M.F. and Carson, C.T. 2013. Multiplex flow cytometry barcoding and antibody arrays identify surface antigen profiles of primary and metastatic colon cancer cell lines. *PloS One,* 8(1), pp.e53015.

Tanskanen, J., Saarela, S., Tankka, S., Kalkkinen, N., Rhen, M., Korhonen, T.K. and Westerlund-Wikstrom, B. 2001. The gaf fimbrial gene cluster of escherichia coli expresses a full-size and a truncated soluble adhesin protein. *Journal of Bacteriology,* 183(2), pp.512-519.

Tao, S.C., Li, Y., Zhou, J., Qian, J., Schnaar, R.L., Zhang, Y., Goldstein, I.J., Zhu, H. and Schneck, J.P. 2008. Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology,* 18(10), pp.761-769.

Tarentino, A.L., Trimble, R.B. and Plummer, T.H. 1989. Enzymatic approaches for studying the structure, synthesis, and processing of glycoproteins. *Methods in Cell Biology,* 32pp.111-139.

Tarnowski, B.I., Spinale, F.G. and Nicholson, J.H. 1991. DAPI as a useful stain for nuclear quantitation. *Biotechnic & Histochemistry,* 66(6), pp.296-302.

Tateno, H., Uchiyama, N., Kuno, A., Togayachi, A., Sato, T., Narimatsu, H. and Hirabayashi, J. 2007. A novel strategy for mammalian cell surface glycome profiling using lectin microarray. *Glycobiology,* 17(10), pp.1138-1146.

Taylor, I. *Forward scatter vs. side scatter* [Online]. Available from: https://www.flowjo.com/learn/flowjo-university/flowjo/getting-started/58 [Accessed 5/16 2017].

Taylor, M.E. and Drickamer, K. 2011. Introduction to glycobiology, 3rd edition. Oxford University Press.

Tielker, D., Hacker, S., Loris, R., Strathmann, M., Wingender, J., Wilhelm, S., Rosenau, F. and Jaeger, K. 2005. Pseudomonas aeruginosa lectin LecB is located in the outer membrane and is involved in biofilm formation. *Microbiology,* 151(5), pp.1313-1323.

Turner, M.W. 2003. The role of mannose-binding lectin in health and disease. *Molecular Immunology,* 40(7), pp.423-429.

Turner, M.W. 2003. The role of mannose-binding lectin in health and disease. *Molecular Immunology,* 40(7), pp.423-429.

Varki, A. Cummings, R. D. Esko, J. D. Freeze, H. H. Stanley, P. Bertozzi, C. R. Hart, G. W. and Etzler, M. E. 2009. Essentials of Glycobiology, 2nd edition. Cold Spring Harbour Laboratory Press, New York.

Wallis, R. 2007. Interactions between mannose-binding lectin and MASPs during complement activation by the lectin pathway. *Immunobiology,* 212(4), pp.289-299.

Wang, P. 2004. Altered sialylation and sialyltransferase expression in gynecologic cancers. *Taiwanese Journal of Obstetrics and Gynecology*, 43(2), pp.53-63.

Ward, A.M., Catto, J. and Hamdy, F. 2001. Prostate specific antigen: Biology, biochemistry and available commercial assays. *Annals of Clinical Biochemistry,* 38(6), pp.633-651.

Wheeler, S.F., Domann, P. and Harvey, D.J. 2009. Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of α (2→ 3)-and α (2→ 6)-isomers. *Rapid Communications in Mass Spectrometry,* 23(2), pp.303-312.

World Health Organisation. *Cancer* [Online]. Available from: http://www.who.int/mediacentre/factsheets/fs297/en/ [Accessed February 15 2017].

Wurm, F.M. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology,* 22(11), pp.1393-1398.

Xu, C. and Ng, D.T. 2015. Glycosylation-directed quality control of protein folding. *Nature Reviews Molecular Cell Biology,* 16(12), pp.742-752.

Yang, R.K. and Sondel, P.M. 2010. Anti-GD2 strategy in the treatment of neuroblastoma. *Drugs of the Future,* 35(8), pp.665.

Yoshida-Moriguchi, T., Yu, L., Stalnaker, S.H., Davis, S., Kunz, S., Madson, M., Oldstone, M.B., Schachter, H., Wells, L. and Campbell, K.P. 2010. O-mannosyl phosphorylation of alpha-dystroglycan is required for laminin binding. *Science (New York, N.Y.),* 327(5961), pp.88-92.

Zhang, L., Luo, S. and Zhang, B. 2016. The use of lectin microarray for assessing glycosylation of therapeutic proteins. *mAbs.* Taylor & Francis, 8(3), pp.524-535.

Zhao, H., Liang, Y., Xu, Z., Wang, L., Zhou, F., Li, Z., Jin, J., Yang, Y., Fang, Z. and Hu, Y. 2008. N-Glycosylation affects the adhesive function of E-Cadherin through modifying the composition of adherens junctions (AJs) in human breast carcinoma cell line MDA-MB-435. *Journal of Cellular Biochemistry,* 104(1), pp.162-175.

Zheng, X., Baker, H., Hancock, W.S., Fawaz, F., McCaman, M. and Pungor, E. 2006. Proteomic analysis for the assessment of different lots of fetal bovine serum as a raw material for cell culture. part IV. application of proteomics to the manufacture of biological drugs. *Biotechnology Progress,* 22(5), pp.1294-1300.

Zimmer, M. 2002. Green fluorescent protein (GFP): Applications, structure, and related photophysical behavior. *Chemical Reviews,* 102(3), pp.759-782.

Appendix A

A.1 LecB and LecB mutant Purifications

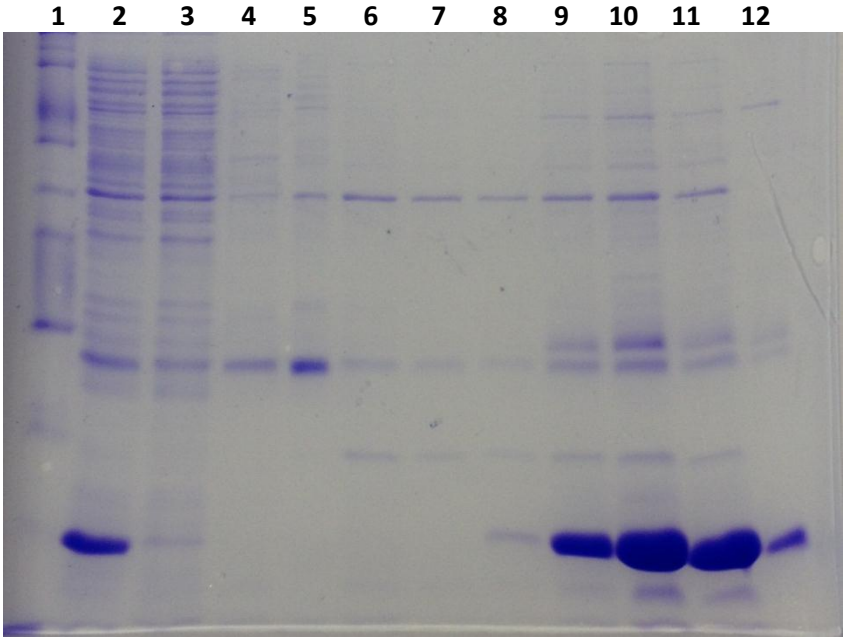


Figure A1.1: LecB purification showing contaminating larger protein bands. Lanes; 1) Broad range protein ladder, 2) Cleared filtered Lysate, 3) Unbound flow through, 4) 20 mM imidazole wash, 5) 80 mM imidazole wash, 6) Elution fraction 1, 7) Elution fraction 2, 8) Elution fraction 3, 9) Elution fraction 4, 10) Elution fraction 5, 11) Elution fraction 6, 12) Elution fraction 7.

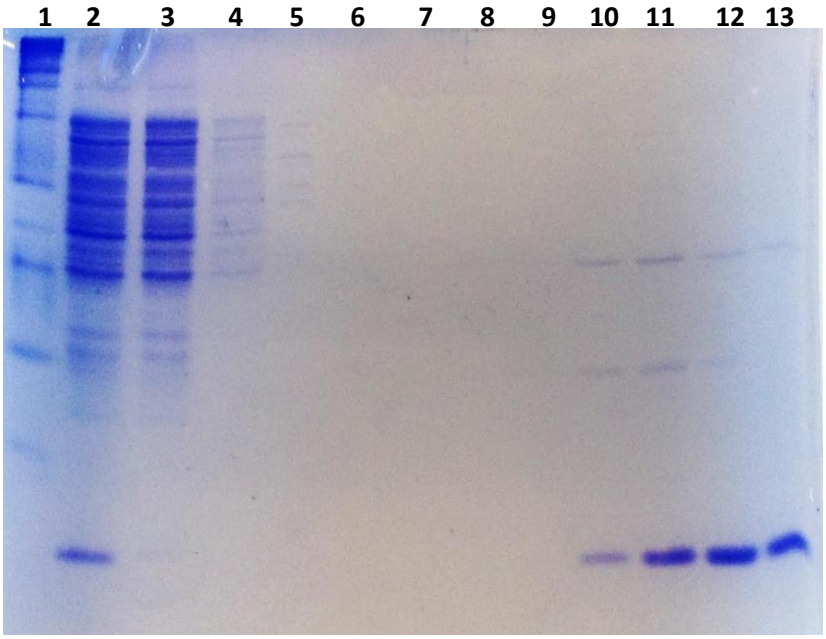


Figure A1.2: LecB S23A;S24A;G25A purification showing minimal contaminating bands with a relative larger size. Lanes; 1) Broad range Protein ladder, 2) Cleared filtered lysate, 3) Unbound flow through, 4) 20 mM imidazole wash, 5) 50 mM imidazole wash, 6) 80 mM imidazole wash, 7-13) Elution fractions 1-7.

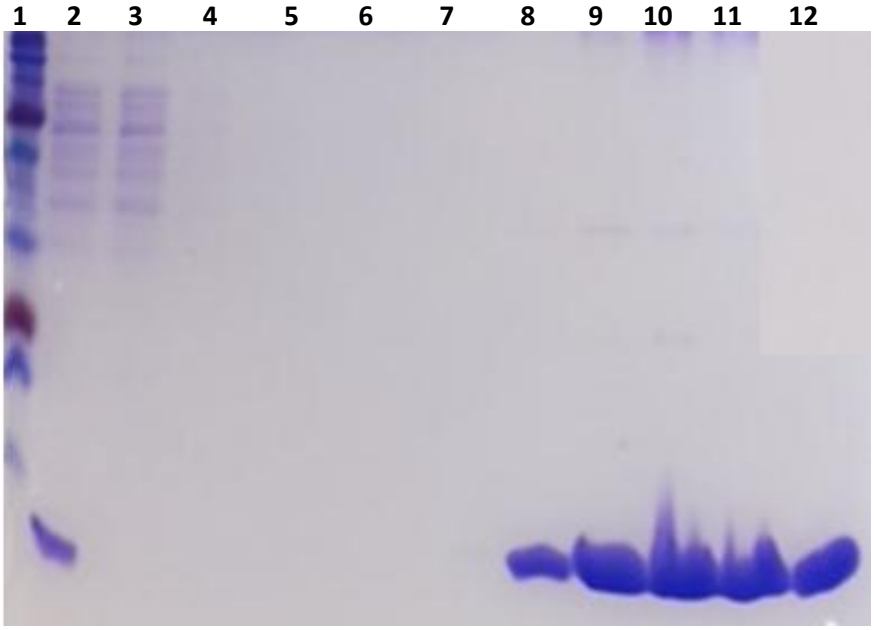


Figure A1.3: LecB S24T purification warped gel with high level of pure protein. Lanes; Broad range Protein ladder, 2) Cleared filtered lysate, 3) Unbound flow through, 4) 20 mM imidazole wash, 5) 50 mM imidazole wash, 6) 80 mM imidazole wash, 7-12) Elution fractions 1-6.

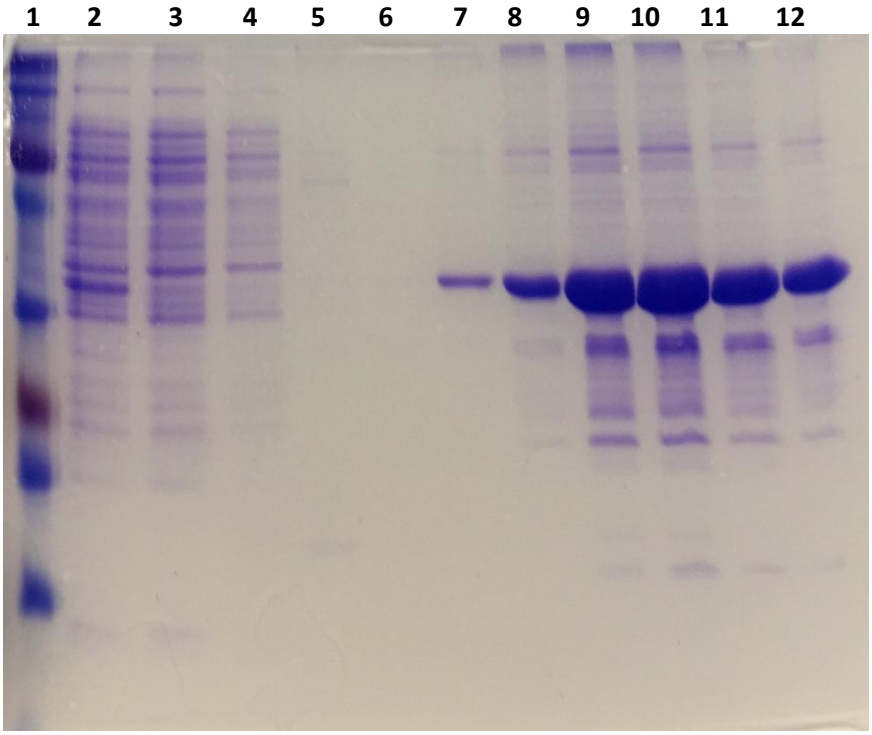


Figure A1.4: EGFP-LecB purification showing numerous contaminating protein bands. Lanes; 1) Broad range Protein ladder, 2) Cleared filtered lysate, 3) Unbound flow through, 4) 20 mM imidazole wash, 5) 50 mM imidazole wash, 6) 80 mM imidazole wash, 7-12) Elution fractions 1-6.

A.2 LecB and LecB mutants Sequence Data (both nucleotide and amino acid)

Amino Acid sequences

LecB wild type:

M R G S H H H H H H G S M A T Q G V F T L P A N T R F G V T A F A N S S G T Q T V N V L V N N E T
A A T F S G Q S T N N A V I G T Q V L N S G S S G K V Q V Q V S V N G R P S D L V S A Q V I L T N E L
N F A L V G S E D G T D N D Y N D A V V V I N W P L G

LecB S24T:

M R G S H H H H H H G S M A T Q G V F T L P A N T R F G V T A F A N S T G T Q T V N V L V N N E T
A A T F S G Q S T N N A V I G T Q V L N S G S S G K V Q V Q V S V N G R P S D L V S A Q V I L T N E L
N F A L V G S E D G T D N D Y N D A V V V I N W P L G

LecB 3K (added lysine):

M R G S H H H H H H G K K K S M A T Q G V F T L P A N T R F G V T A F A N S S G T Q T V N V L V N
N E T A A T F S G Q S T N N A V I G T Q V L N S G S S G K V Q V Q V S V N G R P S D L V S A Q V I L T
N E L N F A L V G S E D G T D N D Y N D A V V V I N W P L G

LecB S23A;S24A;S25A (LecB 3A):

M R G S H H H H H H G S M A T Q G V F T L P A N T R F G V T A F A N A A A T Q T V N V L V N N E T
A A T F S G Q S T N N A V I G T Q V L N S G S S G K V Q V Q V S V N G R P S D L V S A Q V I L T N E L
N F A L V G S E D G T D N D Y N D A V V V I N W P L G

eGFP-LecB:

M R G S H H H H H H G S M S K G E E L F T G V V P I L V E L D G D V N G H K F S V S G E G E G D A T
Y G K L T L K F I C T T G K L P V P W P T L V T T L T Y G V Q C F A R Y P D H M K Q H D F F K S A M P
E G Y V Q E R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I D F K E D G N I L G H K L
E Y N Y N S H N V Y I M A D K Q K N G I K V N F K I R H N I E D G S V V Q L A D H Y Q Q N T P I G D
G P V L L P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G I T H G M D E L Y K G S
M A T Q G V F T L P A N T R F G V T A F A N S S G T Q T V N V L V N N E T A A T F S G Q S T N N A V
I G T Q V L N S G S S G K V Q V Q V S V N G R P S D L V S A Q V I L T N E L N F A L V G S E D G T D N
D Y N D A V V V I N W P L G

eGFP:

M R G S H H H H H H G S M S K G E E L F T G V V P I L V E L D G D V N G H K F S V S G E G E G D A T
Y G K L T L K F I C T T G K L P V P W P T L V T T L T Y G V Q C F A R Y P D H M K Q H D F F K S A M P
E G Y V Q E R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I D F K E D G N I L G H K L
E Y N Y N S H N V Y I M A D K Q K N G I K V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G
P V L L P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G I T H G M D E L Y K G S A C
E L G T P G R P A A K L N

Nucleotide sequences

Note areas of interest are highlighted in specific ways; Red signifies a start codon, Yellow signifies a stop codon, Green signifies a BamHI restriction site, Blue Signifies a HindIII restriction site, Pink signifies an AccI restriction site and all underlined sequences signify where a mutant or insertion has been introduced.

LecB wild type:

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC ATG GCA ACA CAA GGA GTG TTC ACC CTT CCC
GCC AAC ACC CGG TTC GGC GTC ACC GCC TTC GCC AAC TCG TCC GGA ACC CAG ACG GTG AAC GTG CTG
GTC AAC AAC GAG ACG GCC GCG ACC TTC AGC GGG CAA AGC ACC AAT AAC GCC GTC ATC GGC ACC CAG
GTG CTC AAC TCC GGC AGC AGT GGC AAG GTA CAG GTC CAG GTC AGC GTC AAC GGC CGC CCC TCG GAT
CTG GTC TCG GCA CAG GTA ATC CTG ACC AAC GAG CTG AAC TTC GCC CTG GTC GGC TCT GAA GAC GGC
ACC GAC AAC GAC TAC AAC GAC GCC GTC GTG GTG ATC AAC TGG CCG CTC GGC TAG TAA AAG CTT AAT

LecB S24T:

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC ATG GCA ACA CAA GGA GTG TTC ACC CTT CCC
GCC AAC ACC CGG TTC GGC GTC ACC GCC TTC GCC AAC TCG **ACC** GGA ACC CAG ACG GTG AAC GTG CTG
GTC AAC AAC GAG ACG GCC GCG ACC TTC AGC GGG CAA AGC ACC AAT AAC GCC GTC ATC GGC ACC CAG
GTG CTC AAC TCC GGC AGC AGT GGC AAG GTA CAG GTC CAG GTC AGC GTC AAC GGC CGC CCC TCG GAT
CTG GTC TCG GCA CAG GTA ATC CTG ACC AAC GAG CTG AAC TTC GCC CTG GTC GGC TCT GAA GAC GGC
ACC GAC AAC GAC TAC AAC GAC GCC GTC GTG GTG ATC AAC TGG CCG CTC GGC TAG

LecB 3K (added lysine):

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA **AAA AAG AAA** TCC ATG GCA ACA CAA GGA GTG TTC ACC CTT CCC GCC AAC ACC CGG TTC GGC GTC ACC GCC TTC GCC AAC TCG TCC GGA ACC CAG ACG GTG AAC GTG CTG GTC AAC AAC GAG ACG GCC GCG ACC TTC AGC GGG CAA AGC ACC AAT AAC GCC GTC ATC GGC ACC CAG GTG CTC AAC TCC GGC AGC AGT GGC AAG GTA CAG GTC CAG GTC AGC GTC AAC GGC CGC CCC TCG GAT CTG GTC TCG GCA CAG GTA ATC CTG ACC AAC GAG CTG AAC TTC GCC CTG GTC GGC TCT GAA GAC GGC ACC GAC AAC GAC TAC AAC GAC GCC GTC GTG GTG ATC AAC TGG CCG CTC GGC TAG

LecB S23A;S24A;G25A (LecB 3A):

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC ATG GCA ACA CAA GGA GTG TTC ACC CTT CCC GCC AAC ACC CGG TTC GGC GTC ACC GCC TTC GCC AAC **GCG GCC GCA** ACC CAG ACG GTG AAC GTG CTG GTC AAC AAC GAG ACG GCC GCG ACC TTC AGC GGG CAA AGC ACC AAT AAC GCC GTC ATC GGC ACC CAG GTG CTC AAC TCC GGC AGC AGT GGC AAG GTA CAG GTC CAG GTC AGC GTC AAC GGC CGC CCC TCG GAT CTG GTC TCG GCA CAG GTA ATC CTG ACC AAC GAG CTG AAC TTC GCC CTG GTC GGC TCT GAA GAC GGC ACC GAC AAC GAC TAC AAC GAC GCC GTC GTG GTG ATC AAC TGG CCG CTC GGC TAG

eGFP-LecB:

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCT ATG AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC CCA ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTT TCT GTC AGT GGA GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA CTT ACC CTT AAA TTT ATT TGC ACT ACT GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT CTC ACT TAT GGT GTT CAA TGC TTT GCG AGA TAC CCA GAT CAT ATG AAA CAG CAT GAC TTT TTC AAG AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA AGA ACT ATA TTT TTC AAA GAT GAC GGG AAC TAC AAG ACA CGT GCT GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT AGA ATC GAG TTA AAA GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTT GGA CAC AAA TTG GAA TAC AAC TAT AAC TCA CAC AAT GTA TAC ATC ATG GCA GAC AAA CAA AAG AAT GGA ATC AAA GTT AAC TTC AAA ATT AGA CAC AAC ATT GAA GAT GGA AGC GTT CAA CTA GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT GTC CTT TTA CCA GAC AAC CAT TAC CTG TCC ACA CAA TCT GCC CTT TCG AAA GAT CCC AAC GAA AAG AGA GAC CAC ATG GTC CTT CTT GAG TTT GTA ACA GCT GCT GGG ATT ACA CAT GGC ATG GAT GAA CTA TAC AAA GGA TCC ATG GCA ACA CAA GGA GTG TTC ACC CTT CCC GCC AAC ACC CGG TTC GGC GTC ACC GCC TTC GCC AAC TCG TCC GGA ACC CAG ACG GTG AAC GTG CTG GTC AAC AAC GAG ACG GCC GCG ACC TTC AGC GGG CAA AGC ACC AAT AAC GCC GTC ATC GGC ACC CAG GTG CTC AAC TCC GGC AGC AGT GGC AAG GTA CAG GTC CAG GTC AGC GTC AAC GGC CGC CCC TCG GAT CTG GTC TCG GCA CAG GTA ATC CTG ACC AAC GAG CTG AAC TTC GCC CTG GTC GGC TCT GAA GAC GGC ACC GAC AAC GAC TAC AAC GAC GCC GTC GTG GTG ATC AAC TGG CCG CTC GGC TAG

Modified eGFP

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCT ATG AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC CCA ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTT TCT GTC AGT GGA GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA CTT ACC CTT AAA TTT ATT TGC ACT ACT GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT CTC ACT TAT GGT GTT CAA TGC TTT GCG AGA TAC CCA GAT CAT ATG AAA CAG CAT GAC TTT TTC AAG AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA AGA ACT ATA TTT TTC AAA GAT GAC GGG AAC TAC AAG ACA CGT GCT GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT AGA ATC GAG TTA AAA GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTT GGA CAC AAA TTG GAA TAC AAC TAT AAC TCA CAC AAT GTA TAC ATC ATG GCA GAC AAA CAA AAG AAT GGA ATC AAA GTT AAC TTC AAA ATT AGA CAC AAC ATT GAA GAT GGA AGC GTT CAA CTA GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT GTC CTT TTA CCA GAC AAC CAT TAC CTG TCC ACA CAA TCT GCC CTT TCG AAA GAT CCC AAC GAA AAG AGA GAC CAC ATG GTC CTT CTT GAG TTT GTA ACA GCT GCT GGG ATT ACA CAT GGC ATG GAT GAA CTA TAC AAA GGA TCC GCA TGC GAG CTC GGT ACC CCG GGT CGA CCT GCA GCC AAG CTT AAT TAG

A.3 Supplementary ELLA data for recombinant prokaryotic lectin characterisation

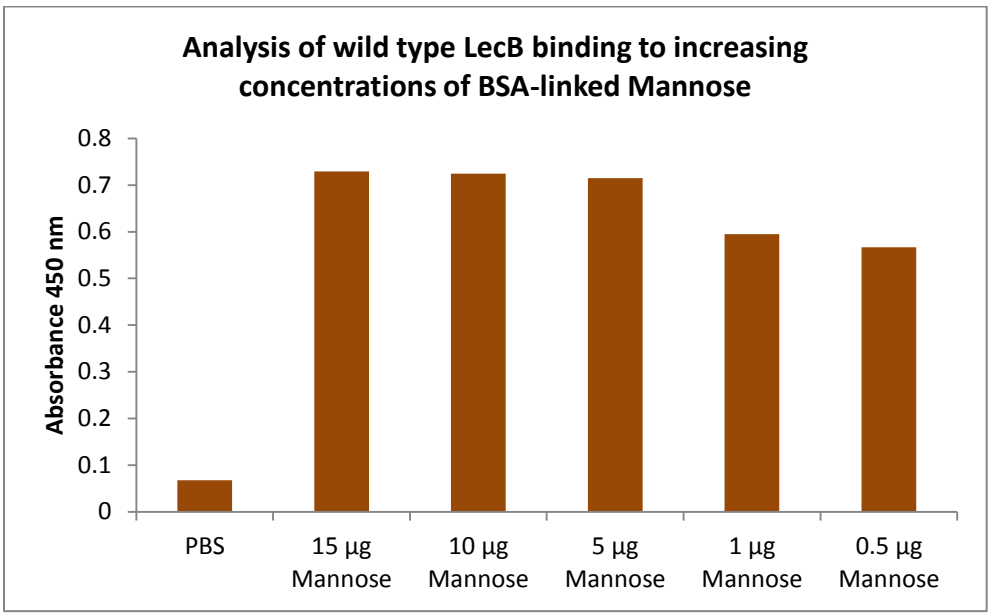


Figure A3.1: ELLA analysis to determine the minimum level of glycoprotein needed for 5 µg/mL wild type LecB to have binding saturation. Increasing concentrations of BSA-linked mannose were added to the plate and probed with 5 µg/mL of LecB wild type.

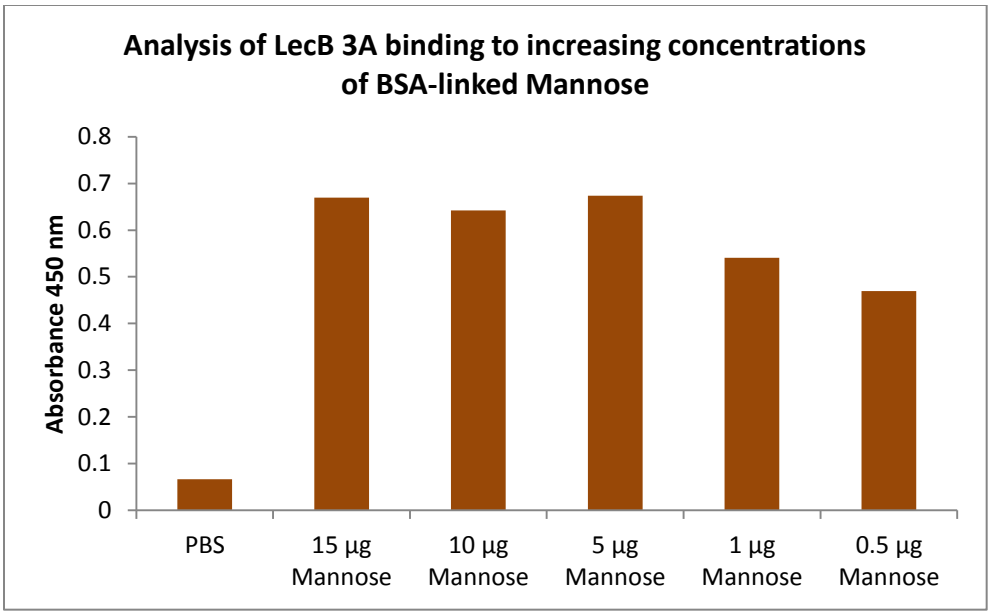


Figure A3.2: ELLA analysis to determine the minimum level of glycoprotein needed for 5 µg/mL LecB 3A to have binding saturation. Increasing concentrations of BSA-linked mannose were added to the plate and probed with 5 µg/mL of LecB 3A.

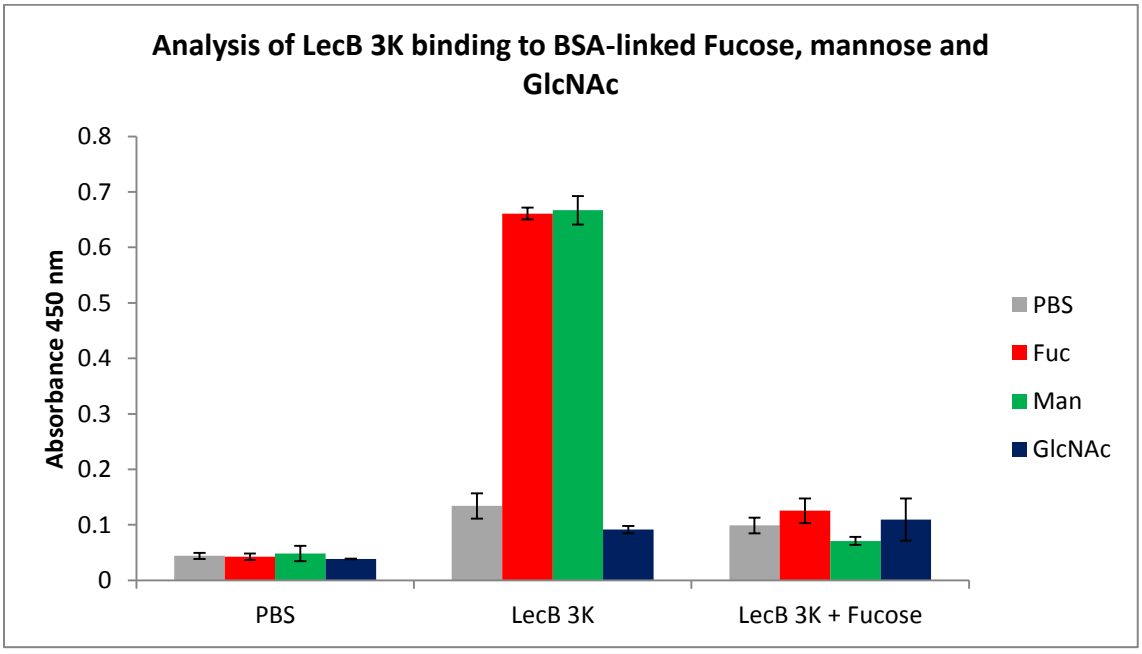
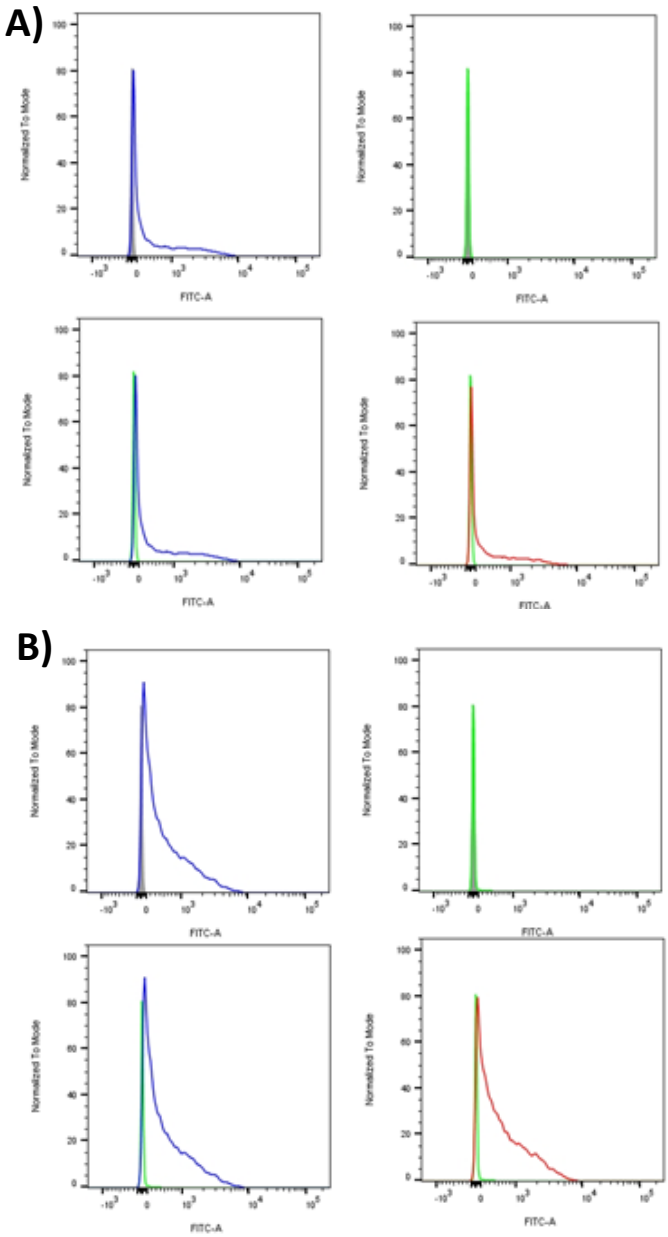


Figure A3.3: Preliminary ELLA analysis of LecB with added lysine probed against BSA-linked fucose, mannose and GlcNAc. Bars on the chart; Grey = PBS, Red = BSA-linked Fucose, Green = BSA-linked Mannose, Navy = BSA-linked GlcNAc.

Appendix B

B.1 Sugar inhibition studies for lectins tested using flow cytometry.



Colour scheme for sugar inhibition histograms;

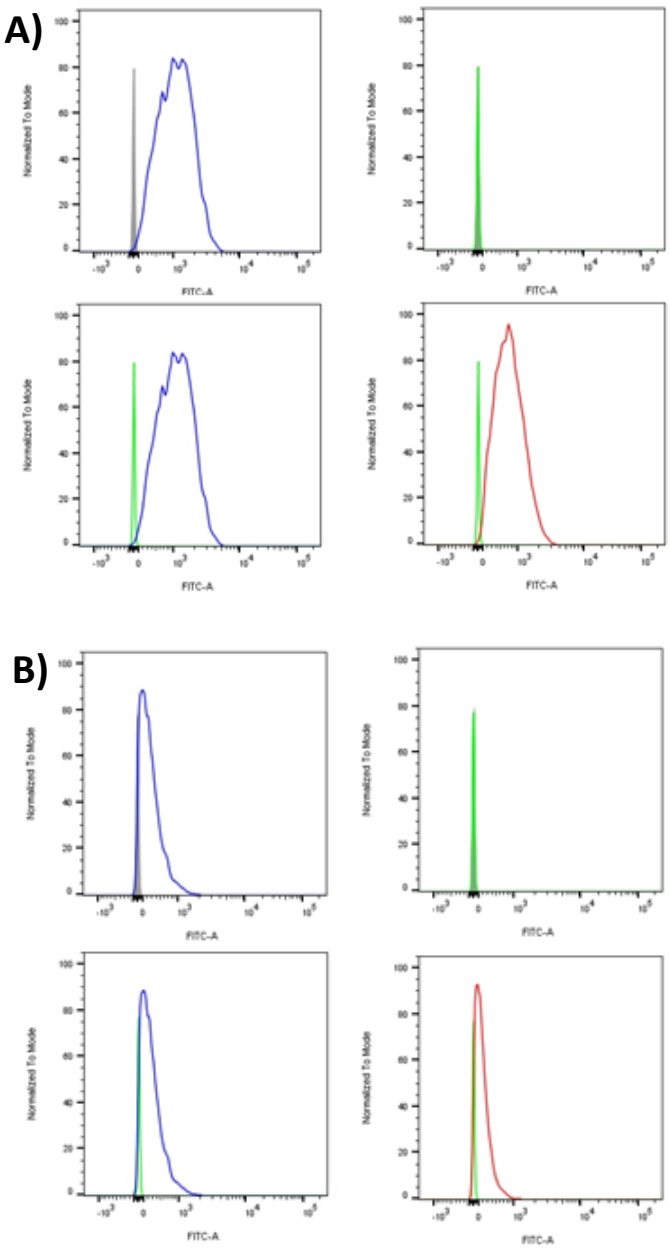
Grey = Unstained Cells

Blue = Cells probed with lectin

Green = Cells probed with lectin pre-incubated with inhibitory sugar

Red = Cells probed with lectin pre-incubated with non-inhibitory sugar

Figure B1.1: SW480 and SW620 cells probed with PNA and PNA pre-incubated with inhibitory and non-inhibitory sugars. A) SW480 cells, B) SW620 cells. Competitive inhibitory sugar use for PNA is 200 mM galactose. The non-inhibitory sugar used for PNA is 200 mM L-fucose.



Colour scheme for sugar inhibition histograms;

Grey = Unstained Cells

Blue = Cells probed with lectin

Green = Cells probed with lectin pre-incubated with inhibitory sugar

Red = Cells probed with lectin pre-incubated with non-inhibitory sugar

Figure B1.2: SW480 and SW620 cells probed with UEA and UEA pre-incubated with inhibitory and non-inhibitory sugars. A) SW480 cells, B) SW620 cells. Competitive inhibitory sugar use for UEA is 200 mM L-Fucose. The non-inhibitory sugar used for UEA is 200 mM D-fucose.

B.2 MTS and fluorofire blue assays to determine lectin cytotoxicity

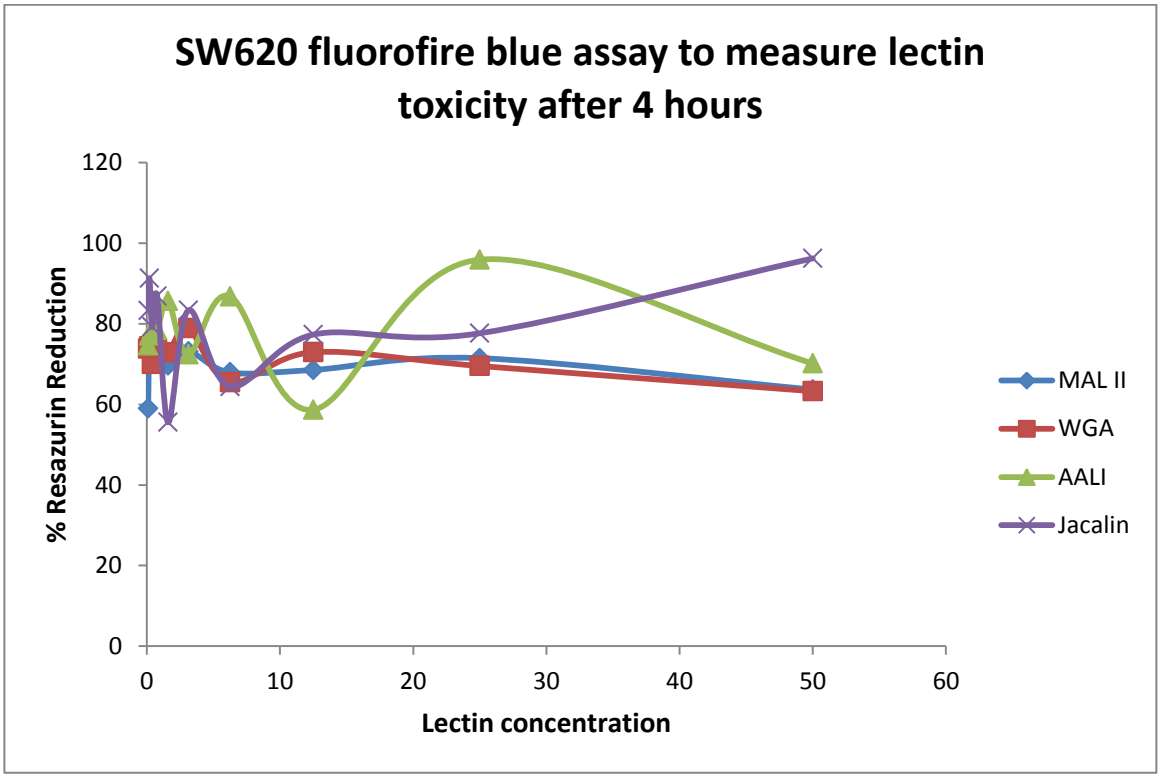


Figure B2.1: Fluorofire blue toxicity assays for lectins MAL II, WGA, AAL I and Jacalin with SW620 cells. The percentage reduction of resazurin is measured and calculated to show the level of metabolically active cells in the sample.

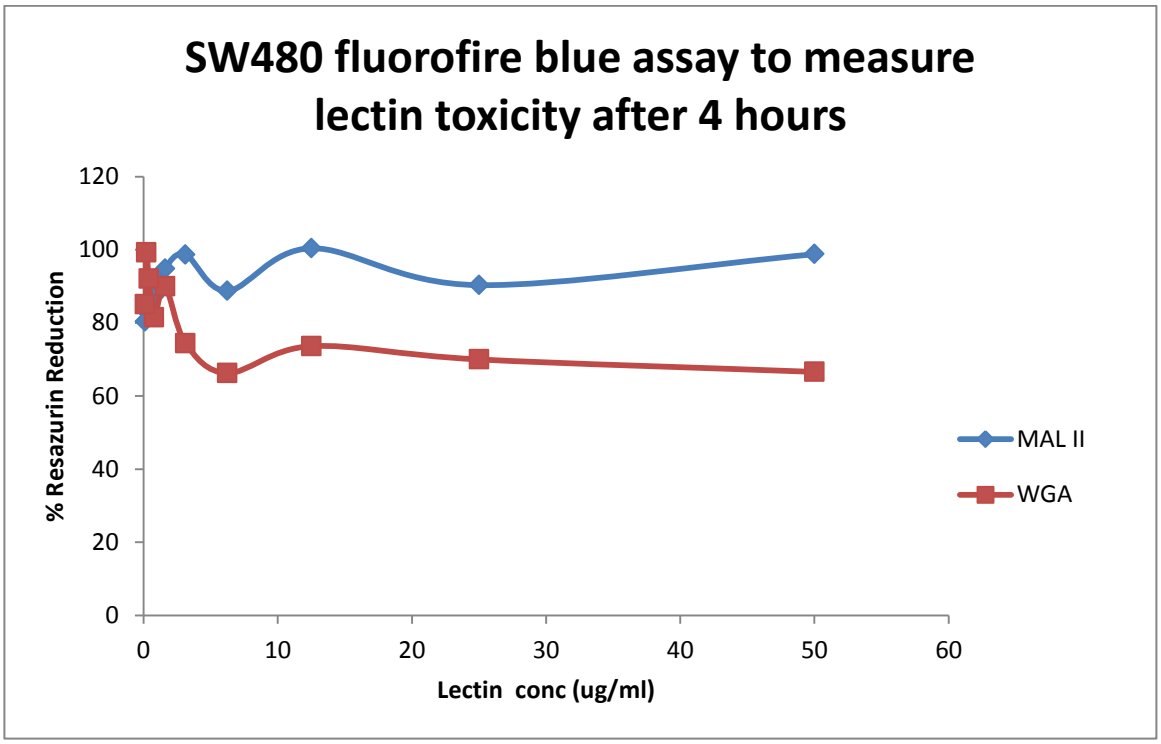


Figure B2.2: Fluorofire blue toxicity assays for lectins MAL II and WGA with SW480 cells. The percentage reduction of resazurin is measured and calculated to show the level of metabolically active cells in the sample.