

1.0 Introduction

1.1 Glycobiology

Glycobiology is defined as the study of biological carbohydrates, and is, after decades of neglect finally emerging as one of the most rapidly growing fields within the natural sciences. The current understanding of this area is considered to be significantly behind the understanding of the other biological information carrying compounds, DNA and proteins, as it is only relatively recently that methodologies have emerged for the large-scale analysis of sugars. The reasons for this are many, and outlined in the following section, but the reasons for the sudden upsurge in the importance and interest in glycobiology are two-fold. Firstly, it has emerged from many studies that carbohydrates (often referred to as glycans) play an important role in a range of recognition events involving cells in a wide variety of tissues and organisms. Secondly, from an industrial perspective, it has emerged that carbohydrates are crucially important in determining the properties of a many biological therapeutics. This study focuses on lectins, a class of glycan binding proteins, and their application on a variety of platforms for use as bio-recognition molecules for glycosylated therapeutics.

1.2 Glycosylation

Glycosylation is a post-translational process that happens to around half of all proteins in eukaryotes (Apweiler *et al.*, 1999). Depending on cell type and organism, the factors that control the events are very different, but it can generally be stated that the most important step is the formation of the sugar amino-acid bond, as it is this step that usually determines the type of carbohydrate unit to be later formed by the cellular machinery (Spiro, 2002).

It was originally thought that glycosylation was a modification that only occurred in eukaryotic organisms, but in the mid-1970s S-layer glycoproteins were found on the

archaebacterium *Halobacterium salinarium*. Glycoproteins have since been identified on many other archaebacteria as well as some eubacteria (Messner 1997). The recent completion of the human genome project has allowed for the identification of the majority of the genes involved in glycosylation, which account for around 2% of the total genes in the human genetic code (Campell and Yarema, 2005). Why an organism would dedicate such a high percentage of its genome to one process becomes apparent when the volume of information that can be carried through carbohydrates is investigated. Unlike nucleic acids and amino acids, glycan assembly is not a template driven process, with no proof-reading enzymes, and carbohydrates have the potential to be assembled in branched structures as well as linearly. As a result there are a colossal 38,016 permutations of three monosaccharides that can theoretically produce a tri-saccharide, compared to 64 possible permutations of four nucleotides in a three-base codon (Laine, 1997). Again, the retention of such of a high percentage of the genome for this process becomes understandable when the benefits of glycans to a proteome are considered. The yeast genome was found to contain ~6,200 genes. When four glycosylation states (no glycans, low, medium and high glycosylation) of the 6,200 encoded proteins are deemed possible, the amount of possible proteins becomes 1.5×10^{15} (Gabius, 2001). Essentially glycans can be considered an organisms mechanism for vastly increasing its proteome, as proteins with the same amino acid sequence but varying glycan structures can have vastly different roles. There are four main types of glycosylation, *N*-linked, *O*-linked, glycosylphosphatidylinositol (GPI)-anchored proteins and glycosaminoglycans. The first three are now briefly discussed. It is important at this juncture to highlight the difference between glycosylation and glycation. Glycosylation is a controlled enzyme driven process, whereas glycation can be defined as the attachment of a sugar molecule to a protein or lipid without the element of enzymatic control, often associated with diabetes (Singh *et al.*, 2001).

1.2.1 N-linked glycosylation

N-linked glycosylation gets its name from the addition of the glycan structure to the nitrogen group of the asparagine within the amino acid triplet which is to be glycosylated. It is initiated by the transfer of a pre-fabricated 14-unit oligosaccharide common to most eukaryotes (Fig 1.1) consisting of *N*-acetylglucosamine, mannose and glucose residues from a dolichol membrane anchor (Fig 1.2) to the Asn located within Asn-X-Ser/Thr triplet where X is any amino acid except proline (Kornfeld and Kornfeld, 1985). This transfer is carried out within the rough endoplasmic reticulum (Fig 1.3) by the enzyme oligosaccharyl transferase (OT). The attached oligosaccharide then undergoes enzymatic trimming within the golgi apparatus by various glycosidases. Other monosaccharide units are then added by further glycosyltransferases often in a tissue or state specific manner (Review by Taylor, 1998).

Although glycosylation is often described as a post-translational event, *N*-linked glycosylation is a co-translational event. As a result, the process is important for the correct folding of the protein, though not crucial, as many studies incorporating *N*-glycosylation inhibitors have shown (Erickson *et al.*, 2007). The calnexin cycle is the quality control system within the ER that sends misfolded glycoproteins to the cytosol for degradation or chaperones correctly folded glycoproteins to the golgi for further processing. The process is reliant on mannose/glucose recognizing proteins that share many features with lectins.

Glycosyltransferases and glycosidases are families of enzymes that cleave or transport sugars which are characterized according to the saccharide unit to which they are specific. The newly synthesized glycoproteins then exit the Golgi and are transported to their final destination. The primary β -glycosylamine bond (GlcNAc) was first identified in ovalbumin (Johansen *et al.*, 1961), and was thereafter found in a vast array of proteins such as enzymes, cell surface receptors, plasma proteins, thyroglobulins, hormones and immunoglobulins. It has also been observed on archaea and, in rare cases, bacteria.

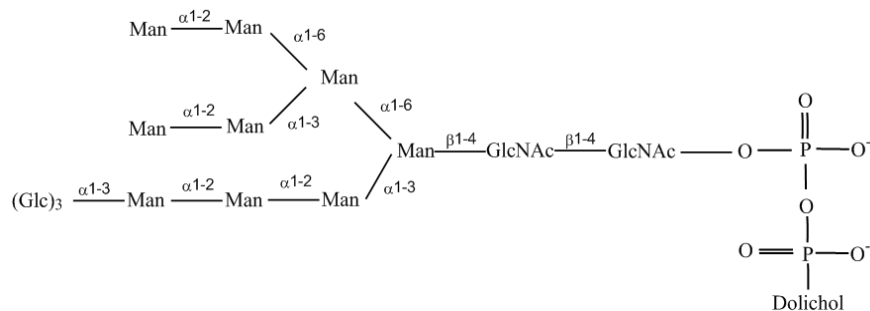


Fig 1.1: The precursor oligosaccharide unit for *N*-glycosylation. Schematic of the of the lipid-linked core oligosaccharide common to all *N*-linked glycans in all eukaryotes (Taylor, 1998)

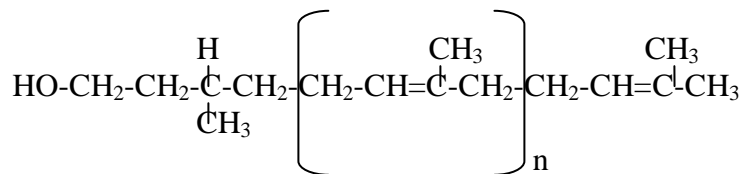


Fig 1.2: Structure of the lipid molecule Dolichol. The precursor oligosaccharide is linked by a pyrophosphoryl group to dolichol. It is a highly hydrophobic molecule, and long enough (75-95 carbon atoms) to span the ER membrane 3-4 times.


Fig 1.3: Biosynthesis of precursor oligosaccharide unit for *N*-linked glycosylation. The process is initiated at the cytosolic face of the ER with the sequential addition of GlcNAc and mannose residues to a dolichol phosphate molecule. Membrane proteins named ‘flippases’ then facilitate the transfer of the oligosaccharide as well as subsequently required free mannose and glucose residues across the ER membrane. The oligosaccharide is then transferred co-translationally to an Asn-residue on the polypeptide through the activity of OT (Image created using ChemBioDraw 11.0).

1.2.2 O-linked glycosylation

O-linked glycosylation produces the second most common glycoform. The carbohydrate residue (GalNAc/Mannose/Galactose/Fucose/Xylose) in *O*-glycans is covalently attached to the peptide backbone via the hydroxyl group of serine, threonine, tyrosine, hydroxyproline, hydroxylysine or another hydroxylated amino acid. In contrast to *N*-glycans, these glycans show a higher degree of structural diversity and do not share a common core structure (Holemann and Seeberger, 2004). As a result, their analysis has proven more difficult than the *N*-linked glycans, with less known about the organisation and structures involved.

The process has been widely characterized for secreted glycoproteins in several yeast strains, where it was shown, like *N*-linked glycosylation, to be initiated in the ER by the transfer of a precursor sugar, in this case mannose, to the β -hydroxyl group of serine or threonine from Dol-P residues by members of the *O*-mannosyl-transferase family. Additional mannose/galactose residues are then added to the primary mannose within the golgi apparatus (Goto, 2007).

In eukaryotes, the most abundant and most widely characterized *O*-glycosylated proteins are the mucins (high molecular weight glycoproteins produced by epithelial cells which are important for the function of mucus membranes) where the variable glycans often total up to 80% of the proteins dry weight. In this case it is GalNAc that acts as the initial sugar and is attached to Ser/Thr residues by a family of ppGalNAc transferases, and then gets modified by further glycosyl-transferases within the golgi (Hang and Bertozzi, 2005)



Complex Mucin-Type
O-linked Glycans



Fig 1.4: Glycosylation of the mucin proteins in eukaryotic organisms. The diagram shows the process through which the complex carbohydrate chains are created within the golgi apparatus using the nomenclature that is explained in Appendix A.

1.2.3 Other glycoforms

Glycosylphosphatidylinositol (GPI)-anchored proteins were first identified in 1988, and have the basic core structure shown in Fig 1.5. It is seen most commonly when eukaryotic cell surface proteins are attached to a phospholipid bilayer through a GPI-anchor. The protein to be anchored is translated with cleavable *N* and *C*-terminal signal sequences. Once the *N*-terminal signal sequence directs the protein to the ER,

the hydrophobic C-terminus is replaced by a pre-formed GPI-anchor (Böhme and Cross, 2002). Variation in this form of glycosylation comes from various substitutions within the GPI-pentasaccharide backbone shown below.

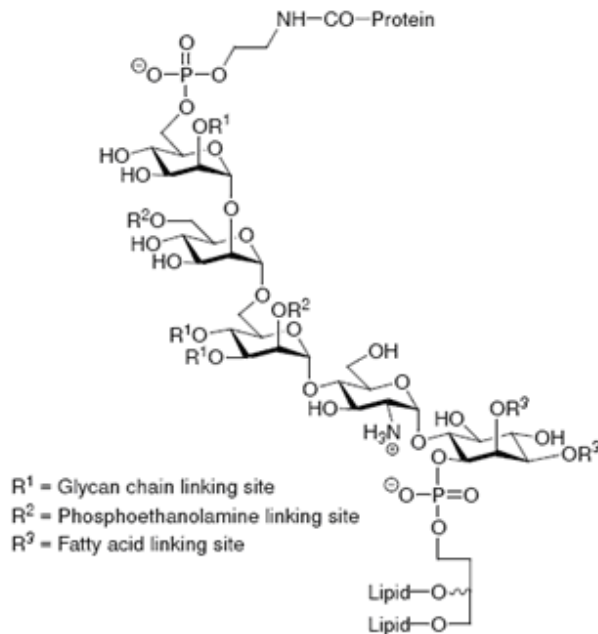


Fig 1.5: Basic core structure of GPI-anchors. (Holemann and Seeberger, 2004).

1.2.4 Bacterial glycosylation

Until the 1970s the idea of protein glycosylation in prokaryotes was widely thought of as impossible, but with more evidence to the contrary being unearthed every year, it is now an accepted fact that bacteria express glycosylated proteins. The first such evidence was the discovery of S-layer glycoproteins in the *Archaea*, which dates back to the 70s, where they were found on halobacteria and thermophilic clostridia (Mescher and Strominger, 1976; Sleytr and Thorne, 1976). In the *Archae* and the Gram-positive bacteria, S-layer glycoproteins account for up to 20% of the total protein output of a cell, and in the case of *Halobacterium halobium*, accounted for 50% of the cell envelope proteins (Mescher *et al.*, 1974). The diversity that has been discovered to date in these glycan structures has far exceeded the range displayed by eukaryotes, with many reviews outlining their functions, structures, synthesis and molecular biology (Reviewed in Schaffer and Messner, 2004).

Non-S-layer glycoproteins have been increasingly the focus of much research, as they have been found to be important in the pathogenesis of many Gram negative bacteria. The flagella of many strains have been found to be extensively glycosylated such as *Campylobacter coli* and *Campylobacter jejuni* (Thibault *et al.*, 2001; Szymanski *et al.*, 2002), *Helicobacter spp.* (Josenhans *et al.*, 1999), and *Treponema pallidum* (Wyss, 1998). Pili and adhesins have also been found to contain glycan residues, as seen in *E. coli*, where the AIDA and TibA adhesion molecule is modified with heptose (Moorman *et al.*, 2002). One adhesin (HMW1) that mediates the attachment of *Haemophilus influenzae* to human endothelial cells has also been found to be modified with galactose, glucose and mannose residues (Grass *et al.*, 2003). The type IV pili of *Neisseria meningitidis* have been shown by carbohydrate labelling to contain both N-linked and O-linked glycans, both of which incorporate galactose (Virji, 1997). The organism *P. aeruginosa* has had the glycan additions to its flagellar subunits extensively characterised, as well as the pathways that control the process. A ~16kb cluster has been identified as the island that is required for the glycosylation process that is specific for a-type flagellins (Arora *et al.*, 2001). It is thought that these surface exposed proteins in pathogenic bacteria display these glycans to mimic their hosts glycan patterns and avoid immune cells.

Some of these proteins are glycosylated using enzymes and pathways normally used in the formation of LOS and LPS structures (e.g. the *E. coli* AIDA and TibA proteins) whilst other have evolved their own glycosylation machinery that are not involved in LPS/LOS formation (e.g. the ‘glycosylation gene islands’ found close to the flagella genes within *Campylobacter*).

This list of various glycosylated proteins shows that to date, many predominantly surface exposed proteins that possess glycan structures have been identified, which has lead to some groups proposing roles in immune response evasion and resistance to proteolytic evasion (Szymanski and Wren, 2005).

Complex glycoprotein based therapeutics are currently expressed in animal cells such as chinese hamster ovary cells. Expression systems based in bacterial cells would be more desirable for regulatory reasons, and as a result glyco-engineering within prokaryotic cells is an emerging industry.

1.3 The importance of biological carbohydrates

So far, the complexity of biological carbohydrates and some mechanisms of protein glycosylation have been briefly discussed, with little emphasis on the biological importance or relevance of these molecules. Like many systems, the significance of a single wheel can only be gauged upon its malfunction, so a number of disease states that have been attributed to problems in the construction of glycan structures are listed below.

Alzheimers disease (Huang *et al.*, 2004), cardiac conditions, respiratory ailments, diabetes, stress, nephropathy (Smith *et al.*, 2006), some auto-immune diseases (Hirschberg, 2001), cystic fibrosis (Xia *et al.*, 2005), arthritis (Tomana *et al.*, 1994) and breast cancer (Dwek *et al.*, 2001) are all well documented to display alterations in normal glycan patterns.

Recently the primary immunodeficiency syndrome (LAD II) was found to occur as a result of two amino acid mutations in the GDP-fucose transporter. The alteration to this fucosyltransferase resulted in the complete lack of fucosylation and also all sialylated and sulfated Lewis derivatives (Hirschberg, 2001). An altered glycosylation pattern is known to occur in cystic fibrosis (CF). The CF transmembrane conductance regulator (CFTR) gene, mutated in CF, has been implicated in the decreased sialylation of glycoconjugates (Dosanjh *et al.*, 1994).

A family of inherited diseases termed congenital disorders of glycosylation (CDGs) are characterized by varying symptoms from stroke and psychomotor retardation to the fragile skin seen in progeroid syndrome. These conditions are caused by inherited defects in genes involved in the biosynthesis of the core pentasaccharide (Fig 1.1) and its transfer to the Asn residue of the glycosylated protein. Due to the huge variety in functionality of *N*-glycans, the symptoms presented by these CDG patients are highly complex, and more CDGs are being identified every year (Aebi and Hennet, 2001).

1.3.1 Aberrant glycosylation in cancer cells.

It has been well established over the last 40 years that the glycosylation pattern in cancerous cells is drastically altered from what is present on a healthy cell, with many of the cancer-specific glycan patterns accounting for the tumour specific antigens used in diagnostics. What is not well understood is the cause of this, and at present the scientific community is not even able to conclude if the aberrant glycans are the cause of, or the result of cancer (Reviewed by Hakamori, 2002).

The connection between glycan-lectin mediated cell signalling and cancer development is a highly complex and involved subject that cannot be easily summarised. There are several reviews that have tackled separate areas within the subject, namely, glycan tumour markers (Fuster and Esko, 2005), *O*-glycans in specific tissue cancers (Brockhausen, 2006), galectins and cancer (van den Brule *et al*, 2004), GPI-glycans and cancer (Filmus, 2001), glycosylation pathways in cancer (Lemaire and Juillerat-Jeanneret, 2006).

For the purposes of illustrating the link between the glycans and cancer sialylation is given as an example. Sialic acid is an important glycan that affects properties such as solubility and efficacy (Section 1.4). Sialic acid presentation on surface glycoproteins is largely controlled by the amount of sialyl-transferases (ST) present in the golgi. Over expression of sialylated *O*-glycans T-antigen, Le^x and Le^a antigens has been widely reported in cancer cells. The presence of these sialic acid residues promotes adherence to many selectins present on the surfaces of epithelial cells. In this way, cancer cells metastasise through endothelial binding. In turn, it was found the regulation of some ST genes is controlled by pro and anti-apoptotic genes and ST expression was found to be dramatically altered in cancer cells. In addition, the picture is further complicated by the expression of different versions of these genes due to splice variants which contributes to the protein specificity and activity, and some ST genes are known to be transcribed using a combination of numerous differently controlled promoters (Reviewed by Wang, 2005)

1.4 Glycobiology and the biopharmaceutical industry

Glycoproteins constitute approximately \$60 billion worth of output from the pharmaceutical industry. The glycans presented on glycoprotein products can have drastic implications in terms of efficacy, activity and immunogenicity. Consequently, carbohydrate heterogeneity has serious implications for regulatory compliance, and often is the cause of additional expensive downstream purification steps. In recent times, the FDA has indicated that sugar moieties on glycoproteins will play a larger role in the future with regards to license approval, which was most recently observed in a request to Genzyme for more information regarding the monitoring and analysis of a glycoprotein product (Myozyme) produced in large-scale cultures (Article in *Nat. Biotechnol* (2008) **26**(6): 592).

This is important to consider as current carbohydrate monitoring techniques are off-line and require a significant amount of sample preparation. In the time it takes to elucidate the glycosylation profile of a product within a fermentation, the process has progressed by a significant length of time. Fermentation time (Hooker *et al.*, 1995) as well as other variable factors such as cell density, nitrogen concentration, pH (Liu *et al.*, 2005), point of induction, and the concentration of heavy metal media components (Kopp *et al.*, 1996; Gu *et al.*, 1997) have all been shown to significantly contribute to glycosylation within a fermentation. Therefore, any improvements in glycan monitoring for the bio-pharmaceutical industry would be very welcome.

As well as being legally obliged to analyse these carbohydrate structures, there are some examples of glycosylation being used to add commercial value to product. The importance of glycosylation to a protein's half-life was displayed when extra *N*-glycosylation sites were engineered into recombinant human EPO being expressed in chinese hamster ovary cells. The end result was the increase in the number of sialic acid groups per molecule from 6-10 to 14-18, which nearly tripled the half-life of the molecule. This new molecule, as well as having the advantage of requiring lower dosage and administration, could also be patented as a new molecule, NESP (Darbepoetin alpha/ Aranesp). A similar strategy was also used to increase the

potency of the molecules Mpl ligand and leptin, which previously lacked any *N*-linked glycans (Elliott *et al.*, 2003).

Any successful alterations to sample stability and production methodology, or formulation improvements, have been shown to provide new IP space, and extend product lifespan, which currently is a major issue for most pharmaceutical companies due to the threat of biosimilar products entering the market as patents expire. Successful examples of this include Pegintron® (Schering Plough) & Pegasys® (Roche), Pegfilgrastim® (PEG-GMCSF) and the Kirin-Amgen v Transkaryotic Therapies infringement decision (EPO manufacture IP- House of Lords 2004) over EPO.

Several initiatives have also exploited technologies that allow for the *in-vitro* glycosylation of drug products, thereby drastically increasing the pharmaco-kinetic properties of a homogenous protein, often with associated improved stability and solubility. Examples include Glythera Inc which modify the HIV-inhibitor G-CSF, and are exploring the modification of Interferon- β 1a.

1.5 Traditional glycoprotein analysis techniques

The highly complex nature of the glycosylation process has already been highlighted. Given that the process is also non-template driven, prediction of glycosylation is currently impossible. As a result, analysis of glycans has concentrated on the post-glycosylation stage, with an emphasis on techniques that require removal of carbohydrate moieties from the molecule that has been glycosylated. Chromatography, mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectrometry techniques have proven most successful. Several reviews have compared the usefulness of MALDI-TOF LC/ESI and MS/MS for the investigation and quantification of carbohydrate moieties (Dell and Morris, 2001; Geyer and Geyer, 2006), with an efficient method recently described by Wada *et al.*, (2007). However mass spectrometry has the disadvantage of not only being an extremely sophisticated technique requiring highly skilled operators, but also being expensive to establish and operate.

Another problem facing glycobiologists is that structural information is sparse due to the lack of crystallographic data. X-ray crystallography relies on the immobilisation of proteins in single crystals, which proves complicated for glycoproteins owing to the considerable intrinsic mobility of carbohydrates (Wyss *et al.*, 1996).

Whether studying industrial glycoproteins or more complex structures such as the extracellular matrices, the glycocalyx and lipopolysaccharide, composed of glycolipids as well as glycosaminoglycans, the field of glycobiology is reaching a major bottleneck in terms of data analysis. A widely known image of the glycocalyx is shown in Fig 1.4, which gives an idea of the complexities facing glycobiologists. Given that this sugar layer is heterogenous, and varies according to cellular health and location, even when tools are developed that can identify the individual glycan moieties present within a complex mixture, the presentation of these results in a comprehensible manner to scientists in other fields will prove a major obstacle. In this way, glycobiology shares many common problems with bioinformatics.

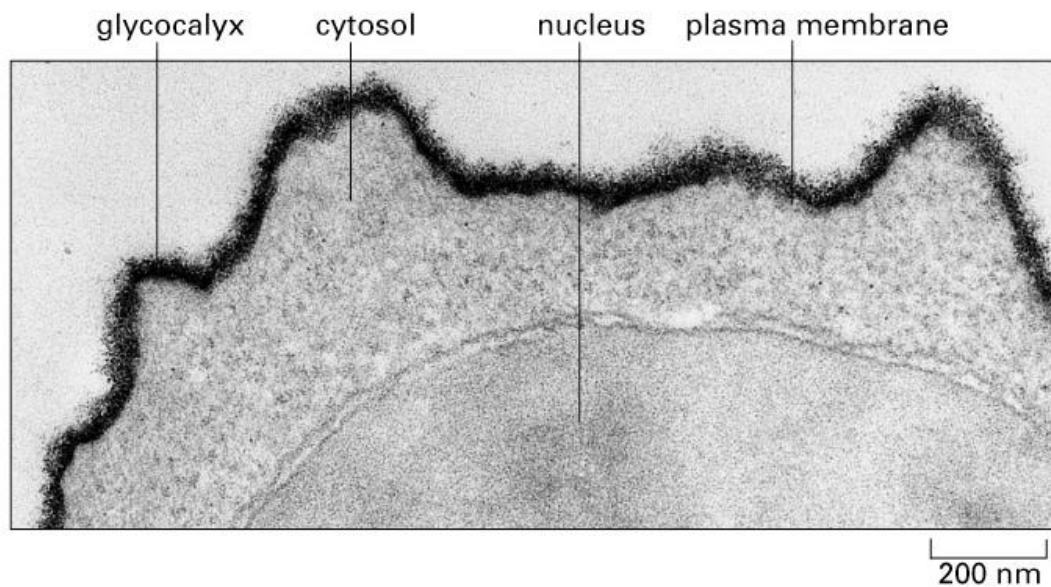


Fig 1.6: Image of the glycan layer surrounding an endothelial cell. The glycocalyx is the dense layer of complex glycans that surrounds the epithelial cell, which has been found to play important roles in the immune system, cell adhesion, fertilization and embryonic development. (Image taken from *Mol.Biol. of the Cell*, 4th Edition. Bruce Alberts et al, Garland Science, N.Y. USA.)

As with analytical procedure, the analysis of a glycan depends on the ability to have a purified homogenous analyte. With this requirement in mind it is appropriate to introduce a family of proteins called lectins. These are carbohydrate binding molecules with no enzymatic activity and are not immunologically derived. They have evolved to recognize specific glycans and are found in a wide variety of organisms, both prokaryotic and eukaryotic. Many regard lectins as the most obvious tool to decipher the glycode, as nature has created so many of these molecules already. To this end, lectins have proven an invaluable tool to chemists and biochemists alike, in that they allow for the separation of glycoproteins from a complex mixture of proteins. However this isolated glycoprotein may still contain a convoluted mixture of highly varied glycoforms which can hinder analysis without further separation.

1.6 Lectins

The first recorded description of a lectin can be dated back to 1888, when Peter Hermann Stillmark of the University of Dorpat (now Tartu, Estonia), described in his doctoral thesis a protein found in plant cell extract that had the ability to agglutinate erythrocytes. These proteins become known as hemagglutinins or phytoagglutinins, as they were generally found in plants. Stillmark isolated one such hemagglutinin from the seeds of *Ricinus communis*, and later called it ricin. Ricin, along with another agglutinin, abrin, were soon commercially available, and became model antigens in very early immunological studies at the turn of the last century.

It would not be until 1936 before it was demonstrated that hemagglutinins were sugar specific, when Concanavalin A (ConA), isolated from jack bean (*Canavalia ensiformis*), was shown to agglutinate cells such as erythrocytes and yeast, and was inhibited by the addition of sucrose. The group (Sumner and Howell, 1936) thus suggested that the agglutination activity was a result of a reaction between the plant protein and sugars that coated the cells surface.

Over the following decades, hemagglutinins became the area of intense interest in relation to the differentiation of different blood types. Many hemagglutinins specific for A, O, B, N and other blood types were discovered, and through inhibition

studies, their respective specific sugars revealed. Through various such discoveries, the term lectin was coined, from the latin *legere*, which means to pick out or to choose (Boyd and Shapleigh, 1954). This term has since become widespread for the description of all sugar specific agglutinins of a non-immune origin, irrespective of source and blood specificity.

Typically, the affinity of lectins for individual carbohydrate units is low, generally in the micro to milli-molar range. Nature however compensates this low affinity by increasing the valency of the carbohydrate-protein interactions (Mulvey *et al.*, 2001). This ‘Velcro™’ effect is achieved through the assembly of individual lectins, each containing a single carbohydrate-binding site. This co-operative binding has the end result of significantly increasing the affinity of a lectin for its target. This technique is not unique to lectins, and can also be seen in filamentous pili, receptors on cell/viral walls and microbial toxins.

With the field of glycobiology rapidly being recognised as critical in the next phase of biological and medical research, it has become clear that the tools for the analysis and evaluation of glycans and carbohydrates are not as well defined as those in the areas of genomics and proteomics. New methodologies and technologies are being designed for application in the expanding glycobiology sector, and amongst these lectins play an enormous role. Created by nature to perform a specific function of adhering to particular sugar residues, these molecules can be exploited in any number of formats to aid biologists to understand the complex field that is glycobiology.

1.7 Lectin families

1.7.1 Legume lectins

There are currently over 100 characterised members of the legume lectin family, mostly purified from seeds of the plants from which they are derived. One of the first lectins to be isolated from the group was ConA, which is an abundant lectin in the jack bean plant. It became the subject of intense study after it was found to agglutinate tumour cells more readily than normal cells (Inbar and Sachs 1969). As a

model lectin, ConA has a number of traits, which are common characteristic of the legume lectins. It has a relatively low molecular mass of 26,500 Da (Min *et al.*, 1992) and its structure is depicted in Fig 1.7.

Most legume lectins consist of two or four monomeric units of between 25-30 kDa each. Each of these typically contains one sugar binding site, as well as a tightly bound Ca^{2+} and a transition metal ion, usually Mg^{2+} . Up to 20% amino acid similarity exists between all legume lectins, with the region involved in the coordination of metal ions being most highly conserved (Ambrosi *et al.*, 2005). The high similarity is not restricted to the primary level, as all legume lectins studied to date have been found to contain the same tertiary structure, which is of two anti-parallel β -sheets, a six-stranded flat “back” and a seven-stranded curved “front”, connected by a five-stranded β -sheet, giving the well known “jellyroll” motif, also referred to as the ‘lectin’ fold.

Their varying specificities can allow for the further classification of this family of lectins into five main groups; mannose, galactose/*N*-acetylgalactosamine, *N*-acetylglucosamine, fucose and sialic acid (Lis and Sharon, 1998).



Fig 1.7: Monomeric structure of the lectin ConA. 3D structure of the legume lectin ConA, with its associated metal ions. Calcium is denoted as a green sphere, and magnesium as a pink sphere. These are located within the sugar binding pocket. Image created using Deepview. (PDB Code: 2CTV)

1.7.2 Cereal lectins

These lectins are commonly found as dimers, with cysteine rich sub-units, which differentiates them from the legume lectins, a family that usually lacks the residue. Wheat germ agglutinin (WGA) is the model lectin from the family and has been well characterized. An image depicting the dimeric structure and the reliance on disulfide binds is given in Fig 1.8.

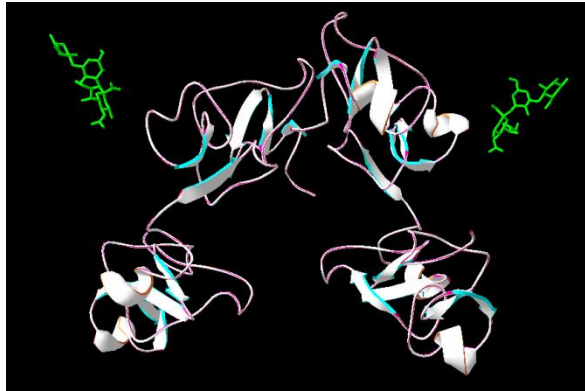


Fig 1.8: Crystal structure of the WGA lectin complexed with lactose. The two WGA monomers can be seen, one on the left, one on the right, each binding to a lactose molecule (green). (PDB Code: 1WGC)

1.7.3 Amaryllidaceae and related family lectins

The bulbs of the Amaryllis, garlic and orchid families contain mannose binding lectins that share a large number of characteristics. The amino acid sequence displays between 80-90% homology, and the lectins are around 12 kDa in size. Members of this family do not require a metal ion for sugar binding, and also display relatively weak binding affinities. A distinguishing feature is the presence of 3 internal repeats of 36 amino acids (Van Damme *et al.*, 1994). An example of a lectin from this family is that from *Galanthus nivalis* (GNA), the structure of which is shown in Fig 1.9.

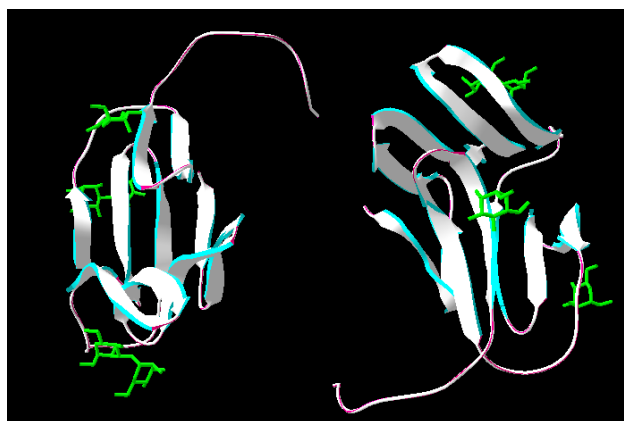


Fig 1.9: Crystal structure of the GNA lectin complexed with mannose α (1-3)methyl-mannose. The two separate monomer structures are clearly visible left and right, with the associated sugars highlighted in green (PDB Code: 1NIV)

The role of plant lectins in nature is unclear; however it is generally believed that they may protect plants against phytopathogenic fungi and insects. This has been shown *in vitro* with the lectins WGA, PNA and SBA inhibiting the growth of plant pathogens of the *Aspergilli* and *Penicilli* families (Barkai-Golan *et al.*, 1978). Several theories also exist which suggest a role in the establishment of symbiotic relationships between leguminous plants and nitrogen-fixing bacteria, which have been reviewed Kijne *et al.*, (1997).

1.7.4 Galectins (formerly S-type lectins)

Galectins are a family of 14 soluble β -D-galactopyranoside recognizing proteins that are found predominantly in mammals, though they have also been reported in other vertebrates and invertebrates (Lis and Sharon, 1998). They have been shown to be important modulators of inflammatory processes, as well as being implicated in having roles in tumour growth and metastasis (Danguy *et al.*, 2002). They are structurally very simple molecules, all exhibiting a highly conserved S-carbohydrate recognition domain (S-CRD). These small subunits (~14 kDa) either form multimeric homodimers, or contain one or two copies of the S-CRD with an accessory region (Rabinovich *et al.*, 2002). Despite little sequence similarity to the legume lectins, the galectins exhibit the same jelly-roll topology (Fig 1.10).

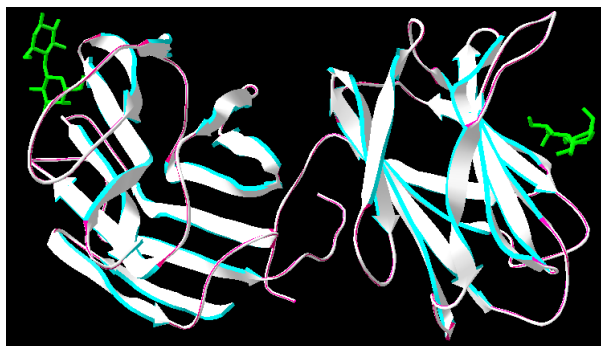


Fig 1.10: Crystal structure of the fungal galectin CGL-2 complexed with lactose. The CGL-2 monomers can be seen at the left and right of the image, each containing one lactose molecule (green) within the binding pocket. (PDB Code: 1ULC)

1.7.5 Complex carbohydrate binding molecules

The aforementioned lectin families are often grouped together and termed the ‘Simple Lectins’. More complex carbohydrate binding molecules are abundant in nature, ranging from viral hemagglutinins, which recognize target carbohydrate moieties on host cell surfaces, to the immunological lectins covered in Table 1.1. There are many immunological lectins as the highly variable lipopolysaccharide layer of bacteria proves an invaluable target for host organisms in the detection and combatting of invading micro-organisms, and provokes responses such as inflammation, antibody formation and abscess formation.

Animal lectins, particularly those in the immunological system have been the subject of more in-depth analysis, and the primary roles of some of the main families are summarized in Table 1.1.

The function of microbial lectins has been easier to elucidate, due to the practical advantages of dealing with prokaryotic organisms compared to eukaryotes. Several specific lectins have been characterised already, with various proposed functions proposed for each (Section 1.7.6).

Table 1.1: Summary of animal lectins

Lectin Type	Role	Target Carbohydrate	Properties
Immunological Lectins			
Ficolins	Complement Pathway	GlcNAc, GalNAc, ManNAc	Four domains: <i>N</i> -terminal, collagen, neck and fibrinogen domains
Pentraxins	Complement Pathway	Galactose, Galactans, Fungal Extracts	Disc-shaped pentamers, Ca dependant
I-Type	Membrane receptors on B-cells	Sialylated moieties	7 extracellular domains, Mediates B-cell responses
F-Box	Innate Immunity	Fucosylated Glycans	Extracellular, characterised by presence of 2 domains
Siglecs	Immunological Regulation	Sialylated moieties	Contain immunoglobulin domains
Housekeeping Lectins			
P-Type	Transmembrane signalling	Man-6-phosphate	Oligomerisation unknown May exhibit 'cluster' effect
M-type	Unfolded protein removal	Mannose	Transmembrane proteins short cytoplasmic tails
L-type	Protein trafficking within ER & Golgi	Various	Common CRD in plant, animal and fungi
Chitinase-like	Development & tissue remodelling	Chitin	Soluble, intracellular and secreted, barrel structure
R-Type	Enzyme Targeting	Varied	Also exists in prokaryotes, Often exist as domains
Heparin-binding	Extra-cellular matrix signalling	Heparin	Members often grouped with other lectins.
Intelectins	Fertilisation and embryogenesis	Gal/Galactofuranose pentoses	Very simple structure, CRD & fibrinogen domain

1.7.6 Bacterial and viral lectins

Carbohydrates have long been known to be specific attachment sites for pathogen recognition (Hooper and Gordon, 2001) and in all of the following examples prokaryotic carbohydrate binding proteins are produced for that purpose. It is known that bacteria possess a significant array of carbohydrate binding proteins, which can fall into three main categories, soluble extra-cellular lectins, carbohydrate recognising domains within larger toxins, and fimbriae/pilli localised lectins.

Studies on microbe recognition of target carbohydrate molecules have a long history, beginning with work on the influenza virus in the early 1950s. Focused research has isolated specific host sialic acid residues that play a crucial role in the complexing of the influenza neuraminidase and hemagglutinin to the host glycans

(Eisen *et al.*, 1997), which has led to the development of several sialic acid analogs as therapeutic pharmaceuticals against the virus (Gubareva *et al.*, 2000).

Two bacterial lectins *Pseudomonas aeruginosa* Lectin-I (PA-IL) and *Photobacterium luminescens* Lectin-I (PL-IL) will be discussed at length later (Section 1.10), but some other well characterised lectins are briefly described here.

There are two reported glycan recognising proteins in *Ralstonia solanacearum*, RS-IL and RS-IIL, which bind to specific carbohydrates present on the primary cell wall of plant cells, and are hypothesised to play a role in the infection of plants by *R. solanacearum* (Kostlanova *et al.*, 2005).

The homologues of RS-IL, *Chromobacterium violaceum* lectin CV-IL (Sudakevitz *et al.*, 2004), and PA-IIL in *Pseudomonas aeruginosa* (Gilboa-Garber, 1982) both have proposed roles in binding to the primary cell wall of plants, and fucosylated glycan structures of epithelial cells respectively. Lectin homologues of PA-IIL and PA-IL (See Section 1.9) are widespread throughout organisms such as *Burkholderia spp.* and *Ralstonia spp.*. In *P. aeruginosa*, these molecules were found to have roles in the formation and maintenance of biofilm structures (Diggle *et al.*, 2006; Sonowane *et al.*, 2006). Highlighted recently was the importance of biofilms for *Ralstonia pickettii* infections (Ryan *et al.*, 2006), and without any reported investigation into the role of lectins within this organism, they remain a target for the development of novel therapeutics.

Although they aren't strictly lectins, there is a similar galactose binding site present in the cholera toxin (CT) and in *Escherichia coli* heat labile enterotoxin (LT), both of which are major virulence factors and are thought to bind to a specific GM1 ganglioside receptor, which is the subject for intense study at present (Holmner *et al.*, 2007).

Another important toxin that contains a glycan recognition site is the pertussis toxin (PT) of the bacterium *Bordetella pertussis*, the causative agent of whooping cough. The toxin, like LT and CT enters the target cell by endocytosis, and is believed to attach to the cell wall through sialic acid containing glycoconjugates (Hazes *et al.*, 1996).

Clostridium botulinum produces a neurotoxin called type C16S progenitor toxin that contains several hemagglutinating subunits. These play an important role in the internalisation of the toxin into target epithelial cells by binding to target glycoconjugates on the cell surface (Nakamura *et al.*, 2007).

The cyanobacterial lectin cyanovirin-N (CVN) is a carbohydrate-binding protein of great interest as it has the potential to bind to the viral envelope protein gp120 of the HIV virus, and thus inhibit viral entry (Bewley *et al.*, 2001). Another inhibitor of viral entry through binding to the same HIV-glycoprotein is the cyanobacterial lectin MVL (Williams *et al.*, 2005).

To date, bacterial derived lectins have not received the same attention as lectins derived from more complex organisms (Fig 1.11). As lectin applications are becoming more complex and diverse (Section 1.8), this pattern should be reversed. This is due to the fact that many non-bacterial derived lectins are themselves glycosylated, which can hinder their usefulness in many applications. This is not an issue with bacterial-derived lectins, and as a result, data obtained from their use in many assay formats have the advantage of not showing discrepancies associated with their own glycans.

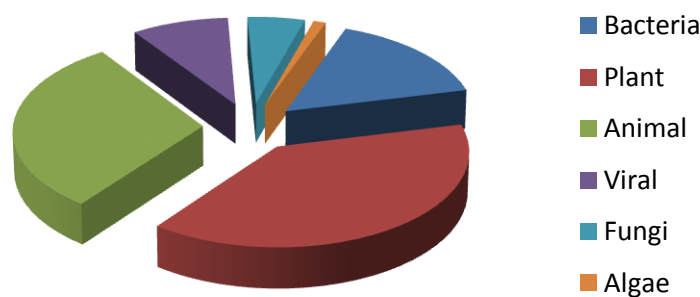


Fig 1.11: Source of available lectin 3D-crystal structures studied to date. The proportion of bacterial and eukaryotic lectins that have been crystallized to date. Bacterial lectins account for only ~15%. (Data source; The lectine database, Centre Nationale de la Recherche Scientifique, <http://www.cermav.cnrs.fr/lectines/>)

1.8 Lectin applications

Lectin carbohydrate interactions are becoming an evermore significant area in glycobiology. The term ‘lectinomics’ has even been coined to encompass the study and application of these varied carbohydrate binding molecules (Gabijs *et al.*,

2001). Some established and emerging applications of these molecules are now discussed.

1.8.1 The enzyme linked lectin assay (ELLA)

Lectin-carbohydrate binding assays are now commonplace, having been cited many times in the literature (McCoy *et al.*, 1983; Rogerieux *et al.*, 1993; Gornik and Lauc 2007). These assays employ enzyme linked solid phase binding assay formats, and are generally done in microplate systems, with either lectin or glycoprotein immobilised onto microtiter plates. These methods can be both quantitative and qualitative, assayed usually by an enzyme linked secondary antibody similar to the indirect ELISA (Engvall and Perlman 1971; Van Weemen and Schuurs 1971) for antigen detection. The ELLA provides an accurate biochemical assay with high sensitivity and applicability to a variety of targets to determine both the sugar binding affinity of the probing lectin, but also the constitution of a complex glycan present either in a complex solution or as a conjugate to a glycolipid or glycoprotein. It also provides a more rapid, cost-effective and simpler alternative to the MS and HPLC alternative methods used for carbohydrate profiling (Section 1.5).

The assay has not only been used for the assaying of glycoproteins of interest, as labelled lectins have already been employed for applications including the investigation of biofilm formation and EPS composition in bacteria (Neu and Lawrence, 1997; Leriche *et al.*, 2000), in eukaryotes (Holloway and Cowen, 1997) and also in cyanobacterium (Kawaguchi and Decho, 2000). Multi-species communities within biofilms have also been investigated using derivatised lectins (Neu and Lawrence, 2001)

Though it shares many characteristics with the ELISA technique, its use worldwide has not proven as commonplace as the original immunological based method (78,342 articles with ELISA vs 89 articles with ELLA in *Title* or *Abstract* fields of PubMed between 1971, when the ELISA was first published, and 2009). As discussed in Section 1.8.5, one of the main problems with any analytical process is the separation of the analyte from contaminant molecules. Immunological assays have the advantage of being able to select solely for the molecule of interest,

whereas the lectin will select any molecule that contains a carbohydrate residue to which it shows specificity. To overcome this problem some groups have attempted to combine the antibodies and lectins in ELISA formats, although the process becomes complicated through the knowledge that many immunoglobulins are themselves glycosylated.

1.8.2 Lectin arrays

Microarrays allow for a multitude of interactions to be analysed simultaneously. Lectins are currently being utilised on a commercial microarray platform available from Procognia Ltd. The system was able to successfully evaluate the glycosylation states of ovalbumin, bovine submaxillary mucin, and porcine gastric mucin (Pilobello *et al.*, 2005). A similar system named the 'QProteomeTM Lectin Array' from Qiagen Ltd was used to successfully investigate the glycan structures present on porcine thyroglobulin, Tamm-Horsfall glycoprotein, and recombinant human erythropoietin (Rosenfeld *et al.*, 2007).

Like the ELLA technique, the requirement for highly skilled operators and equipment is eliminated by employing lectin arrays, so it again provides a useful alternative to MS and HPLC techniques (Section 1.5). However, as the lectins are immobilised on the array surface, either the analyte needs to be labelled to observe a binding event, or a labelled antibody to the analyte must be employed to confirm binding. This limits the applicability of the assay.

One of the main obstacles to any lectin array is the potential loss in protein activity due to attachment of the lectin to the array surface. The immobilisation strategy employed can greatly affect to what extent this will be a factor. The three main techniques are physical, covalent and bioaffinity immobilisation. Physical immobilisation relies on the natural hydrophobic, polar and ionic forces that will form between the lectin and the array surface, with a resulting heterogenous and randomly orientated protein layer. Covalent immobilisation involves the irreversible interaction between exposed amino acids on the protein and modified solid supports, which can result in random orientations depending on the frequency of a specific amino acid throughout the protein surface. The final strategy utilises bioaffinity

immobilisation, and produces a reproducible and homogenous array surface, which can also have the added advantage of being re-usable.

Currently lectin arrays have proved problematic for a number of reasons. Firstly that they are comprised of lectins from several different sources, ultimately leads to a platform which consists of molecules that require different reaction conditions. For example the commonly used plant lectin MAII (*Maackia amurensis* agglutinin – II) requires treated (and hence expensive) blocking agent such as oxidised BSA, as it is very sensitive to contaminant sialoglycoproteins (Kim *et al.*, 2008). Certain plant lectin families require metal co-factors, which are not necessary for others. The same problem exists for optimum pH and temperature conditions for each lectin family.

1.8.3 Lectin affinity chromatography (LAC)

Lectin affinity chromatography is a powerful tool in the arsenal of glycobiologists. It is so effective because it combines a powerful separation principle (chromatography) with a selective and biologically significant bio-recognition phenomenon (lectin-sugar affinity). One of the first examples of molecules being immobilised onto activated-Sepharose for the purposed of separating proteins was the immobilisation of L-tyrosine-D-tryptophan for the purification of carboxypeptidaseA (Cuatrecasas *et al.*, 1968). It was quickly observed that lectins would prove useful in the separation of sugar containing molecules.

Whether within a fermentation environment or in a complex biological sample such as blood or serum, lectin based methods remain the obvious means to characterise glycosylation changes on a glycoprotein. The prevalence of sialic acid, an industrially important terminal glycan (Section 1.4), has been investigated through lectin affinity chromatography. The sialic acid binding lectin *Sambucus nigra* agglutinin was immobilised on a column using simple covalent techniques, and the concentration of sialic acid containing glycoproteins in a sample could be evaluated (Qiu and Regneir, 2005).

Multi-lectin affinity chromatography (M-LAC) has been used to selectively enrich glycoprotein fractions within serum (Yang and Hancock 2004). This was done using

the broad spectrum lectins ConA, WGA and jacalin lectin. It was proposed that this method would have potential in plotting healthy versus diseased state glycoproteins, and therefore could be used as a diagnostic tool in the detection of diseases that contained carbohydrate biomarkers, such as many cancers and liver abnormalities (Section 1.3). Another report of M-LAC being used as a diagnostic tool was the categorisation of *N*-linked moieties of prostate specific antigen (PSA) into prostate carcinoma or benign prostatic hyperplasia by M-LAC, due to an increase in the GlcNAc β (1-4) Man glycan (Sumi *et al.*, 1999).

Lectin affinity chromatography columns are also being developed as downstream process tools for the purification of glycoproteins and other glycan containing targets. For example, using the Human influenza A/Puerto Rico/8/34 virus produced in Madin Darby canine kidney cells as a model, two glycoprotein antigens can be utilized to purify up to 97% of the target virus. In this case the particular virus encodes two glycoproteins on its viral coat, which are located on the viral envelope. Immobilization of the galactophilic lectins from *Erythrina cristagalli* and *Euonymus europaeus* allowed for the capture of the target virus (Opitz *et al.*, 2006). The same technique could be used as a polishing step for the removal of antigenic glycans, or the enrichment of biologically significant sialylated glycoproteins.

The potential for this technology within the biopharmaceutical and diagnostics industry is tremendous, and slowly being recognised by both industries. The limiting factor for the technology is the lack of well-characterised lectins that can be produced cost-effectively.

1.8.4 Lectin delivery molecules

An even more radical application of lectins is their utilization to improve the delivery and targeting of compounds to their target site. This idea exploits the fact that the majority of cells in the body are covered in glycosylated surface proteins. The idea of ‘bio-adhesive’ drug delivery systems has been in circulation since the 1980s, with several examples of commercially available bio-adhesive drugs available. The objective of these molecules are twofold; to increase the residence time of the therapeutic at the biological target, and to concentrate the administered

therapeutic at only the specific biological target. Of specific interest were mucosal specific drug carriers, and the original adhesion molecules were polymer based. However, they had a number of drawbacks, most significantly the issue of non-specific interactions. With the advent of lectinomics, some groups have investigated the application of lectins as an alternative to the traditional polymer based drug carriers. And some examples of studied carriers are given in Table 1.2.

An important consideration when using lectin carrier molecules is that they should not illicit any immune response, as that could render a loss or reduction in the effect of the therapeutic agent. Some plant lectins have been found to be highly immunogenic (e.g. the mistletoe lectins MLI, MLII, and MLIII, wheat germ agglutinin WGA, the tomato lectin LEA, and the *Ulex europaeus* lectin UEA-I) after high levels of specific IgA were found in mice after an extensive immunological survey (Lavelle *et al.*, 2000). Antibodies specific to the lectins WGA, soybean lectin, and peanut agglutinin were also found in human serum (Tchernychev and Wilchek., 1996).

Table 1.2: Lectins studied for the purpose of drug targeting.

Lectin	Specificity	Identified Target	Reference
ConA (Jack Bean)	Mannose	GI-Tract	Lehr and Pusztai, 1995
GNA (Snowdrop)	Mannose		
PHA (Kidney Bean)	Core Structure		
MLI, II & III (Mistletoe)	Gal/GalNAc	GI Tract	Haltner <i>et al.</i> , 1997
UDA (Nettle)	GlcNAc		
WGA (Wheat Germ)	GlcNAc		
MPA (Osage-Orange)	GalNAc	Alveolae	Kasper <i>et al.</i> , 1993
RCA (Castor Bean)	Gal	Alveolae	Kasper <i>et al.</i> , 1993
BSI-B ₄ (Griffonia)	Gal	Nasal Mucosa	Takata <i>et al.</i> , 2000
WGA (Wheat germ)	GlcNAc	Blood/Brain Barrier	Fischer and Kissel, 2001

1.8.5 Eukaryotic versus prokaryotic Lectins

One of the main barriers to the wide-spread use of lectins in research and industry is cost. The most commonly used lectins are the plant lectins, as they are best characterized, and these are usually purified from source. It is widely understood that plant lectins when derived from source can themselves often be glycosylated. One associated problem is lectin cross-reactivity, which was highlighted in a recent study where the broadly specific lectin ConA was found to bind to the plant lectins MAA, SNA and UEA-I when they were plant derived, but not to their *E. coli* produced equivalents (Hsu *et al.*, 2008).

Production of recombinant plant lectins, which has to date mainly been attempted in the bacterium *Escherichia coli* and the yeast *Pichia pastoris* has had limited success, as many plant lectins have been found to form insoluble aggregates when expressed in simpler systems, e.g. SBA in *E. coli* (Adar *et al.*, 1997), and the seed lectin of *Dolichus biflorus*, in which 20% of the lectin was found in the soluble fraction (Chao *et al.*, 1994). PHA was found to be hyper glycosylated in *P. pastoris* (Raemaekers *et al.*, 1999). One recent study has attempted to overcome solubility problems by producing Gst-fusion lectins in *E. coli* for immobilisation onto glass slides (Hsu *et al.*, 2008), however the effect of introducing such affinity tags to the lectin was not established.

Prokaryotic derived recombinant lectins are not glycosylated and hence are more amenable to production and purification for large scale profiling studies. The most commonly used and commercially successfully biorecognition molecules include protein A and streptavidin, both prokaryotic derived and amenable to expression in *E. coli*. Many legume lectins are abundant at source, and as a result facilitate high yields of protein from minimal source material, but this is not the case with all lectin families. In contrast it is not uncommon for recombinant proteins expressed in bacterial systems to achieve up to g/L quantities of purified product.

1.8.6 Lectin mutagenesis

It is generally accepted that lectins have evolved from simple mannose/glucose binding molecules into galactose binders, and finally to sialic acid/fucose binders. Given that the presence of more complex glycans only exist in higher order organisms, this appears to make sense, and the theory is backed up by the presence of trace features of galactose binding molecules within the sialic acid binding plant lectins SNA (*Sambucus nigra*), MAL agglutinin (*Maackia amurensis*) and the galectins.

The mutagenesis of lectins to alter their sugar specificity is not a novel field of lectinomics. However, the disadvantage for researchers concentrating on working with eukaryotic lectins is that manipulations of carbohydrate recognition domains is not as straightforward as it would have been had they been concentrating on prokaryotic sources.

The potential of using bacterial lectins was highlighted in a study where the amino acid sequence of the *P. aeruginosa* lectin PA-IIL was mutated within the sugar-binding loop to the sequence of its homologues in *R. solanacearum* and *C. violaceum*, resulting in a change in its specificity from fucose to mannose (Adam *et al.*, 2007).

One eukaryotic lectin that was shown to be particularly amenable to expression in *E. coli* was β -ricin, which is known to be β -Gal specific. Using error-prone PCR, the specificity of this was successfully altered to α (2-6)sialic acid, to which the parent molecule had no affinity (Yabe *et al.*, 2007). Given the industrial significance of sialic binding lectins, lectins created in this way could have great potential in terms of IP and patentable material.

1.9 The galactophilic lectin PA-IL

Having summarized briefly some of the key applications for lectins, it is clear they share many obstacles to their more widespread use. Many of the problems currently associated with using plant lectins may be overcome by the use of a core bacterial lectin with its carbohydrate recognition domain altered to recognize a variety of varying carbohydrates. A good candidate is the previously characterized PA-IL lectin.

1.9.1 Discovery

The gram-negative opportunistic human pathogen *P. aeruginosa* was found to produce an agglutinin (Gilboa-Garber, 1972a), which was later found to be galactophilic in nature (Gilboa-Garber, 1972b). Initial purification of the then named ‘bacterial hemagglutinin’ was done via a sequence of steps involving the removal of nucleic acids by streptomycin sulfate, precipitation of other proteins by heating to 70°C and acidification, fractionation by ammonium sulfate followed by dialysis and another heating (Gilboa-Garber, 1972a). This crude method was then improved upon, by using affinity chromatography using the carbohydrate matrix Sepharose-4B (Gilboa-Garber *et al.*, 1972). This method produced high yields of purified PA-IL, which provided a basis for the many other PA-IL studies that would follow over the years. The gene that encodes PA-IL was identified, and has since been termed *lecA* as well as *pa-IL* (Avichezer *et al.*, 1992). In this study PA-IL will be the predominant name for the protein, although in some cases *LecA* will be used. Another lectin, PA-III or *LecB*, was later identified in the same organism (Gilboa-Garber, 1982), which has been shown to have an affinity sugars containing fucose, and is not featured in this study.

1.9.2 Sequence analysis of the *lecA* gene

The *lecA* gene which encodes PA-IL is 369 bp in length, with a Shine-Dalgarno sequence located 9 bp upstream from the start codon (Avichezer *et al.*, 1992). Excluding the initiating methionine the molecular weight of the PA-IL monomer is 12,758 Da, and the protein contains a hydrophilic C-terminal domain (Fig 1.16) but no signal peptide.

The genome of the entomopathogenic organism *Photorhabdus luminescens* was sequenced and a homologue to the gene *lecA* identified (Duchaud *et al.*, 2003). Using the identified *P. aeruginosa* PA-IL 121 amino acid sequence (Avichezer *et al.*, 1992) as a query against the *P. luminescens* genome reveals a 369bp open reading frame (ORF; contig 2483371 to 2483739, in reference NC005126, updated on 20 Jan 2009) encoding a PA-IL like protein comprising 121 amino acids (excluding the initiating methionine) displaying 32% identity and 46% similarity to PA-IL (Fig 1.12). This gene, named *plu2096* due to its position on the genome, to my knowledge has not been studied by other groups, and after extensive literature searches, it was found mentioned once in a study that investigated by 2-D gel electrophoresis the proteome of *P. luminescens* (Turlin *et al.*, 2006), where its predicted protein product was found in the extra-cellular fraction that of a *P. luminescens* stationary phase culture.

It is also known that PA-IL: is produced only in stationary phase cultures, as it dependant on RpoS, a sigma factor that is stationary phase dependant (Winzer *et al.*, 2000). PA-IL expression is controlled by the lux box type element together with RpoS consensus sequences, which are contained upstream of the promoter region. The gene *plu2096* has previously been described as encoding a PA-IL like lectin (Duchaud *et al.*, 2003). This protein shows striking similarity to PA-IL in the sugar binding and calcium associated domains (Fig 1.12). This gene isn't located close to the same lux box-type elements or RpoS sequences, suggesting that transcription is control led by another mechanism.

```

PA-IL: -MAWKGEVLANNBAGQVTSIIYNFGDMITIVAAGWASYGFTQKWGPQGDRHPDQGLICHDAFCGALVMKICNSGTI : 76
PL-IL: MSDMSGSSVPANAENGKSTGLILKQGDITISVVAHGWMVYGFNDNVEWAAPDGPVPNN---PQPSSIATLVAKIAN-KKF : 73
      W G V A N E G T 6I GD I36VA GW YG D P1 LV KI N

      *
PA-IL: PWNTGLERWVAENNVQCATLTIYNDVPGTYGNNSGSEFSVNIQKDS----- : 122
PL-IL: AINGVILHKTVE--VDGEIILLFNDVPGTFGDNSEFEQVEVIIIESRYSPLK : 122
      6 G6 P V G 6 165NDVPGT5G1NSG F V 6

```

Fig 1.12: Sequence alignment of PA-IL and PL-IL. Alignment created using ClustalW and edited using GeneDoc (Section 2.28). Homology is represented by black shading, areas highlighted with a blue box representing those amino acids in PA-IL that are involved in orientation of the calcium ion, and those in green identified as directly contacting the galactose molecule (Cioci *et al.*, 2003).

1.9.2 The structure of PA-IL

The PA-IL lectin has been crystallised in a number of different studies, as a monomer (Liu *et al.*, 2002), as a tetramer in the calcium free state (Cioci *et al.*, 2003), complexed with galactose (Cioci *et al.*, 2003) and complexed with Gal α (1-3)Gal β (1-4)Glc (Blanchard *et al.*, 2008). It is known that it oligomerises into a tetramer, a feature in common with many lectins. The four identical subunits are designated A-D (Fig 1.13) When the quaternary structure is examined closely it is clear that each monomer makes contact with two other subunit monomers. Each monomer is approx 47 Å by 29 Å by 25 Å, and they create a tetramer measuring 85 Å by 53 Å by 25 Å. The A-D interface is predominantly hydrophobic (Fig 1.15), with a core region of approx 3,000 Å³ which is contributed to by the hydrophobic amino acid side groups of Val29, Ala30, Ala31, Gly32, Trp33, Gly43, Pro44, Gly80, Leu81 and Phe82. The hydrophilic C-terminus was found to be involved in tetramerisation as the A terminus interacts with the larger of the β -sheets on the B monomer (Cioci *et al.*, 2003) seen in Fig 1.16..

Two cysteine residues are located in close proximity to each other on the molecule, Cys 59 and Cys 64, (Fig1.17) however these are unlikely to form a disulphide bridge due to their distance apart, and because PA-IL is primarily found in the cytoplasm (Glick *et al.*, 1983), a highly reducing environment that is not conducive to the formation of cysteine-cysteine bonds.

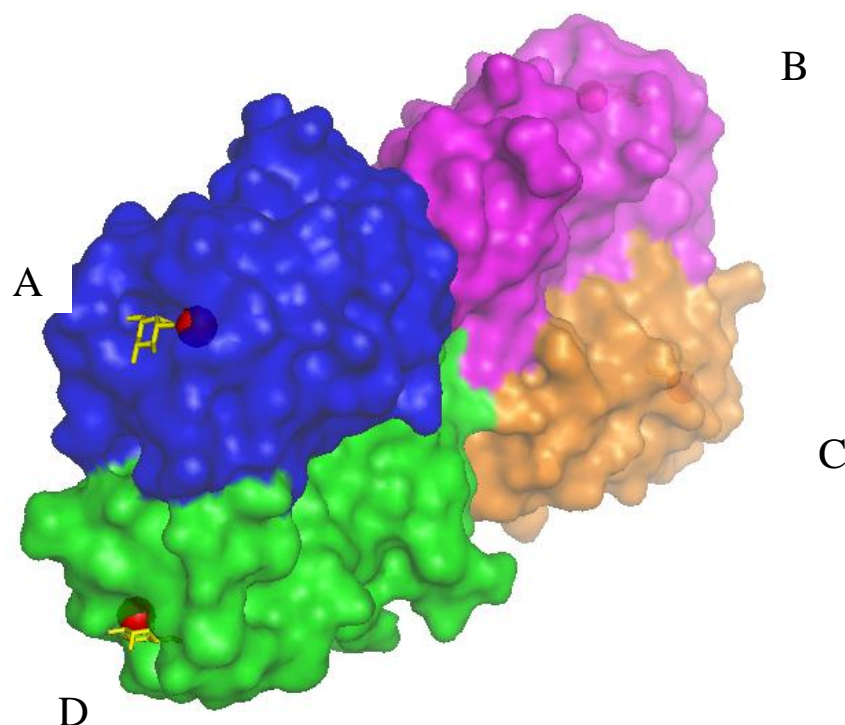


Fig 1.13: Crystal structure of PA-IL complexed with galactose. An image of the crystal structure of four individual PA-IL sub-units forming a tetrameric structure. Calcium ions present in the sugar binding pocket are represented by red spheres, with galactose highlighted in yellow (PDB Code: 1OKO). Image generated using PyMol (Section 2.28).

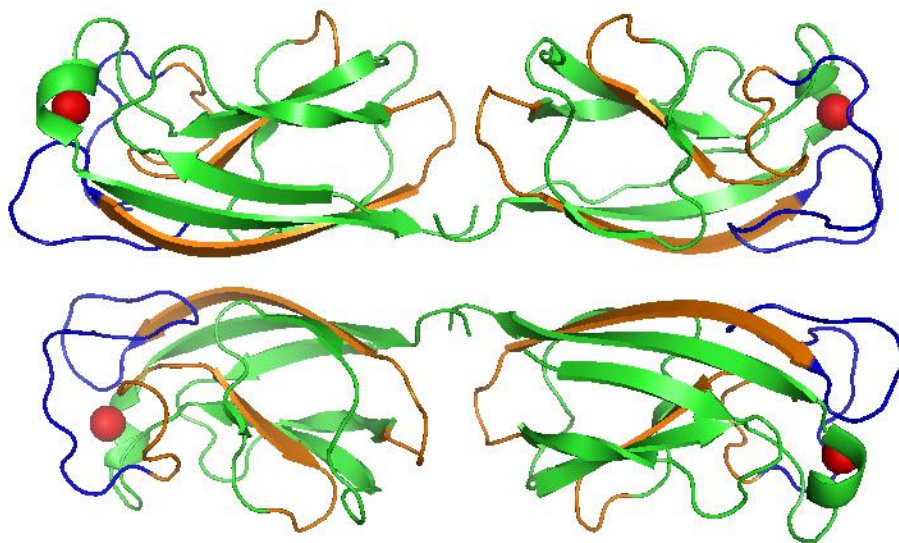


Fig 1.14: Distribution of hydrophilic and hydrophobic regions on PA-IL. Ribbon diagram of the PA-IL tetramer. Residues corresponding to hydrophobic amino acids are represented in blue, with hydrophilic amino acids in orange. The position of the essential calcium ion is indicated with a red sphere. Image generated using PyMol (Section 2.28).

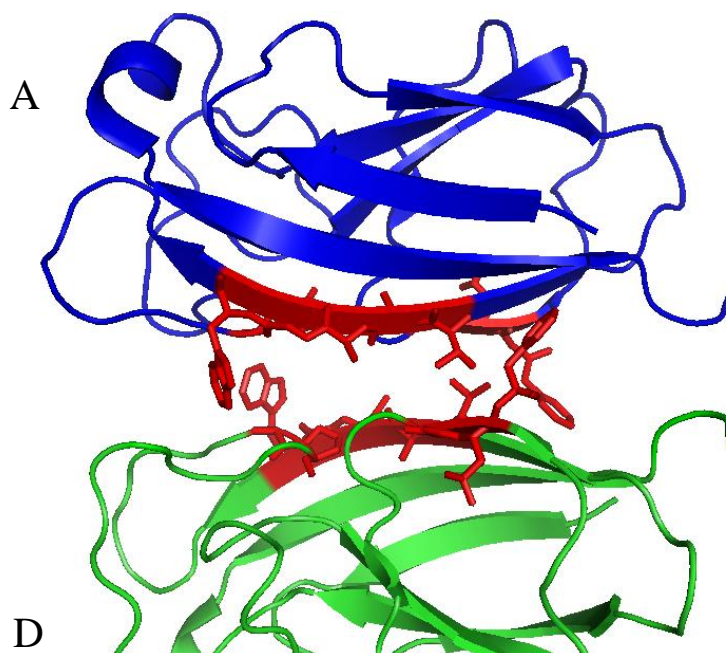


Fig 1.15: The A-D Interface of PA-IL. Ribbon diagram of the A-D interfacial region in PA-IL, with subunit A in blue and subunit D in green. The residues Val29, Ala30, Ala31, Gly32, Trp33, Gly43, Pro44, Gly80, Leu81 and Phe82 are highlighted in red as they contribute to a region of hydrophobicity. Image generated using PyMol (Section 2.28).

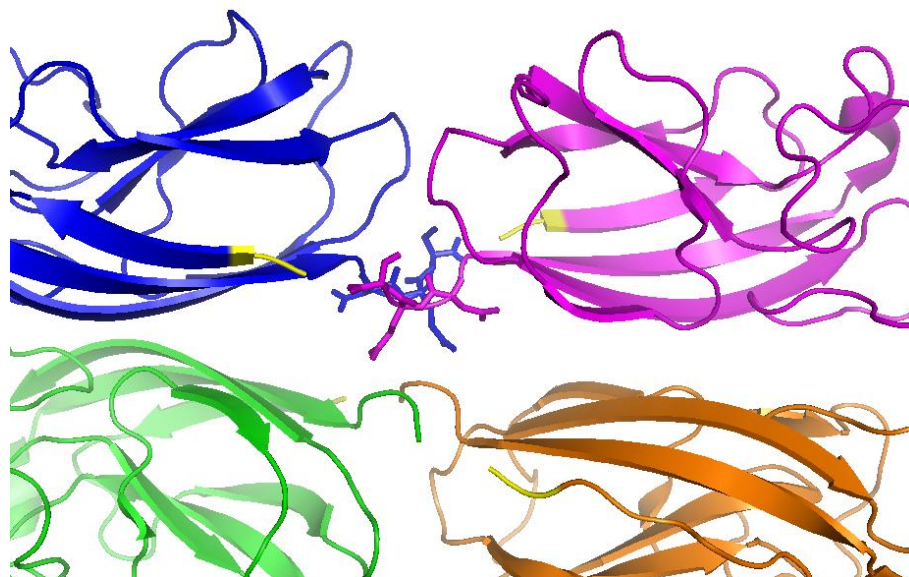


Fig 1.16: The C-termini of PA-IL. Image of the N- and C-terminal regions in the tetrameric PA-IL structure. The N-terminus is highlighted in yellow, with the C-terminal hydrophilic side chains of Asp119, Gln120 and Ser121 in monomers A and B represented by stick figures. Image generated using PyMol (Section 2.28).

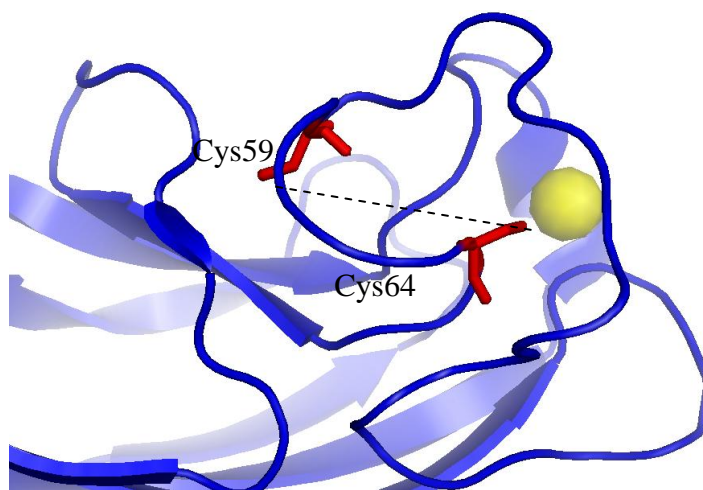


Fig 1.17: The location of cysteine residues on PA-IL. The distance between the two sulphur atoms is 11.49 Å. Image generated using PyMol (Section 2.28).

1.9.3 PA-IL affinity

PA-IL has previously been shown to display medium range affinity for D-galactose (Garber *et al.*, 1992) with a K_a value of $3.4 \times 10^4 \text{ M}^{-1}$ from their conducted equilibrium study. From hemagglutination inhibition studies carried out it is clear that the lectin has a much higher affinity for more complex sugars in the order Phenyl β Gal > p-nitrophenyl-Phenyl β Gal > Melibiose > Stachyose > Raffinose >>> Galactose (Chen *et al.*, 1998). Further to this enzyme linked lectin assays (ELLA) (Kirkeby *et al.*, 2006) and carbohydrate microarrays (Fig 1.18) have shown that the specific α Gal linkages which the lectin prefers are Gal α (1-3)Gal and Gal α (1-2)Gal. This affinity has already been found to be sensitive to temperature, with its ability to agglutinate RBCs diminished at higher temperatures (Gilboa-Garber and Sudakevitz, 1999). This has been disputed however, with the opposite effect described in a separate study (Kirkeby and Moe, 2005). The affinity was shown *in vivo* where much weaker binding of PA-IL to epithelial cells was seen compared to PA-IIL, which prefers higher temperatures (Mewe *et al.*, 2005).

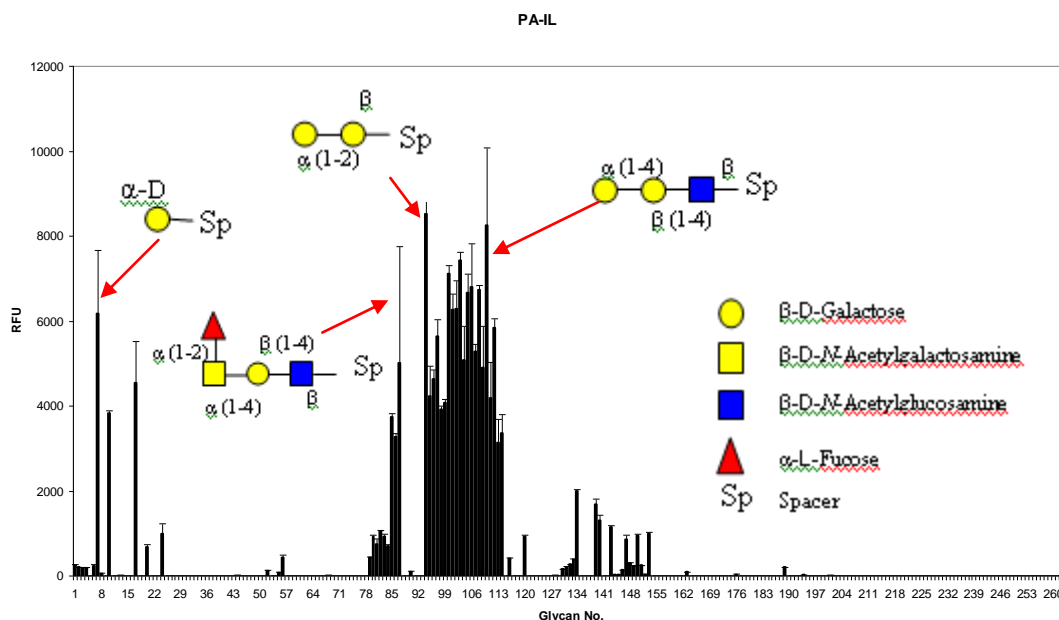


Fig 1.18: Glycan Binding Profile of PA-IL. Carried out by measuring the affinity of labelled PA-IL to biotinylated glycans immobilised on streptavidin coated plates. The structures of some of the positive reacting glycans have been inserted (Carried out by the Consortium for Functional Glycomics (CFG) on behalf of Anne Imberty).

Table 1.3: Summary of identified PA-IL targets.

Target Type	Identified Target	Reference
Burkitts Lymphoma cells	Glycolipid antigen Gb3/CD77 Gal α (1-4) Gal β (1-4)Glc β -Cer	Blanchard <i>et al.</i> , 2008
BSA glyco-conjugates	Gal α (1-3)Gal > Gal α (1-4)Gal > Gal α (1-2)Gal	Kirkeby and Moe, 2005
Mouse endothelial cells	Gal α (1-3)Gal β (1-4)GlcNAc	Kirkeby <i>et al.</i> , 2006
Free dissacharides	Human Blood Group P ^k Gal α (1-4)Gal	Chen <i>et al.</i> , 1998
Mink capillaries & sero-mucinous glands in lung	Unknown	Kirkeby <i>et al.</i> , 2007
Human B and P ^k blood-groups	Gal α (1-3)[Fuc α (1-2)]Gal; Gal α (1-4)Gal	Sudakevitz <i>et al.</i> , 1996

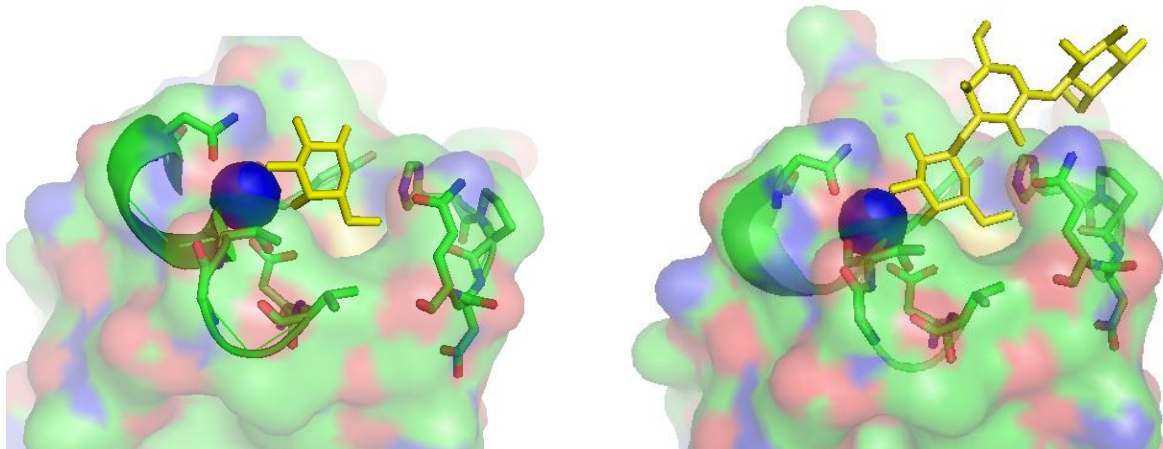


Fig 1.19: A 3D Structure of PA-IL sugar Binding Site complexed with galactose and Gal α (1-3)Gal β (1-4)Glc. Image of the sugar-binding pocket of PA-IL with carbohydrate highlighted in yellow. The location of a calcium ion is depicted with a blue sphere (PDB Codes 1OKO and 2VXJ respectively; Cioci *et al.*, 2003; Blanchard *et al.*, 2008). Image generated using PyMol (Section 2.28).

It has been determined exactly which residues within the PA-IL protein play a direct role in sugar binding from crystallographic studies, with a calcium ion identified as having a crucial role in the mediation of the sugar binding (Cioci *et al.*, 2003). This characteristic is common with many C-type lectins, however, unlike most lectins, PA-IL has a very narrow specificity, namely for α -linked Gal structures. The position of the calcium ion in relation to the sugar binding pocket can be seen in Fig 1.19.

Like many other C-type lectins, the galactose recognition is mediated by the presence of a calcium ion (Gilboa-Garber *et al.*, 2000). Lectin binding curves in reactions lacking the ion have been shown to display a binding curve with a low K_D , while the addition of Ca^{2+} dramatically increases this value. The addition of Mg^{2+} did not have a similar effect on binding kinetics (Kirkeby and Moe 2005).

1.9.4 PA-IL and pathogenicity

P. aeruginosa is found in the regular flora in healthy adults, however its presence in cystic fibrosis and critically ill patients can result in a dramatically increased death rate, and particularly in the case of cystic fibrosis patients, it remains the most important pathogen associated with chronic airway infection (Bajolet-Laudinat *et al.*, 1994). There are a number of proposed mechanisms through which PA-IL plays a role in the pathogenicity of the organism, which will be outlined here.

It has already been shown that PA-IL expression is controlled by the RhlR and LasR quorum sensing regulatory systems (Winzer *et al.*, 2000; Schuster *et al.*, 2003) in *P. aeruginosa* with PA-IL with transcription strictly limited to the stationary phase of growth (Diggle *et al.*, 2002). It is estimated that ~4% of the genes within the organism are regulated by these systems, many of which encode virulence factors (Whitely *et al.*, 1999).

PA-IL is often described as an adhesin. These are molecules produced by microorganisms to overcome the complicated innate and acquired defence mechanisms of eukaryotic cells by binding to the cell surfaces. The tissues and cell types the molecule has been found to attach to are summarised in Table 1.3.

Using the mink as an animal model for human *P. aeruginosa* infection, PA-IL has been found to bind to certain sections of the lung, in particular the seromucinous glands in the submucosa of the large bronchi, capillaries in the alveolar walls, and blood vessels forming the vasa vasorum around the larger vessels, and in the pancreas, the epithelium in the excretory ducts, as well as the pancreatic capillaries (Kirkeby *et al.*, 2007).

Several other *P. aeruginosa* carbohydrate binding proteins have already been identified and investigated which including flagellin and flagellar cap protein FliD, which recognises mucin oligosaccharides, and the pili adhesins (Scharfman *et al.*, 2001), such as type IV pili that bind asialo GM1 and GM2 glycolipids (Hahn, 1997). PA-IL has also been shown to inhibit the ciliary beat, which is an important mechanism through which the host combats bacterial colonization. It was found to have a dose dependent inhibitory effect on the activity of ciliated epithelial cells, and

consequently the ciliary beat frequency (CBF). The inhibition wasn't seen when the same cells were treated with *P aeruginosa* LPS (Bajolet-Laudinat *et al.*, 1994). These results were seen in a later study also, although PA-IL addition was found to initially increase CBF, and was also found to have a lesser CBF-reducing effect than PA-IIL (Mewe *et al.*, 2005).

When *P aeruginosa* is found in the intestinal tract of critically ill patients, there is an associated 70% increase in death rate (Marshall *et al.*, 1993). As a result investigations into the role of the organism's lectins in the gastrointestinal tract have shown that PA-IL results in alterations in epithelial barrier function, which could be prevented with treatment of GalNAc, a binder of PA-IL (Laughlin *et al.*, 2000).

The importance of the lectins PA-IL and PL-IIL can be deduced from the interest of commercial drug companies in the inhibition of these molecules. The synthetic carbohydrate production company Glycomimetics Inc, have licensed a compound called GMI-1051, which inhibits the sugar-specific activity of the PA-IL and PL-IIL lectins of *P. aeruginosa*, and is a useful tool in cases where traditional antibiotic therapies prove ineffective against the pathogen. This field of carbohydrate inhibitory molecules can only increase in importance with more pathogenic bacterial lectins being characterised, and also with the increase in understanding of the importance of selectins in the fields of cancer biology.

1.9.5 Other PA-IL associated functions

Biofilms can be found associated with many types of micro-organisms, and are described as communities of microbes encapsulated within a self-made polymeric matrix, and are usually attached to either a living or an inert surface. PA-IL has been shown to play an important role in the formation of *P. aeruginosa* biofilms in nutrient-depleted environments, as a PA-IL mutant showed inhibited biofilm formation, while an over-expressing mutant showed enhanced biofilm coverage. Similarly, the addition of IPTG, to which PA-IL has a high affinity, successfully dispersed a mature biofilm (Diggle *et al.*, 2006). A contradictory study has also shown that PA-IL doesn't mediate biofilm formation through regulation of membrane proteins, as a *lecA* mutant showed no differences in membrane protein profiling compared to wild type *P. aeruginosa* (Sonawane *et al.*, 2006).

The organism *P. aeruginosa* synthesises several quorum-sensing molecules including N-acetyl homoserine lactones (AHLs). It has been found that the central hydrophobic pocket formed upon the construction of the lectin tetramer has the capacity to bind these AHLs (Boteva *et al.*, 2005)

P. aeruginosa utilises the type II secretory pathway to export a number of proteins including alkaline protease, protease IV and LasB. A PA-IIL but not a PA-IL mutant, was found to significantly reduce the caseinolytic activity of PAK cells. This was found to be due to the loss of protease IV activity (Sonowane *et al.*, 2006). As a result, a role for PA-IL in the proteolytic activity of the organism has been discounted.

1.10 Project aims and objectives

The primary objective of this research project was to clone the gene for *Photorhabdus luminescens* lectin PL-IL which was identified as a homologue to the *Pseudomonas aeruginosa* lectin PA-IL. Upon the cloning and expression of the lectin, it would then be investigated at a structural and functional level.

As much of this characterization has already been carried out on the PA-IL protein, this protein was also cloned, expressed and purified, for use as an internal control within experiments.

Following successful expression and purification of both lectins, large quantities of recombinant protein products would be available for traditional physical and functional methods to characterize the lectins. As well as the traditional methods, newly emerging techniques such as mass spectrometry and adaptations of the existing methods would be utilized. Such studies would be the first on the PL-IL protein.

PA-IL would also be mutated within the sugar binding site to confirm if the sugar specificity of these lectins could be manipulated. This could ultimately lead to the creation of a uniform recombinant lectin microarray with each well containing the same core lectin molecule, but each with a variety of different glycan specificities.

Finally attempts at elucidating the role of the lectin within the organism *P. luminescens* would be made, and any information obtained would be useful in the study of bacterial host interactions, as the bacterium is currently a model organism in the field of mutualistic relationships.

2.0 Materials and Methods

2.1 Bacterial strains, primers and plasmids

The bacterial strains, plasmids and primers used in this study are listed in Tables 2.1, 2.2 and 2.3 respectively. Selected plasmid maps are shown in Fig 2.1 to 2.5.

Table 2.1 Bacterial strains

Strain	Genotype	Features	Source
<i>Escherichia coli</i>			
JM109	F' traD36, <i>proAB</i> ⁺ <i>lacI</i> ^q , Δ <i>lacZ</i> M15 <i>endA1 recA1 hsdR17</i> (r _k ⁻ , m _k ⁺) <i>mcrA</i> <i>supE44</i> λ - <i>gyrA96 relA1</i> Δ (<i>lacproAB</i>) <i>thi-1</i>	All purpose cloning Strain Produces stable plasmid DNA	Sigma
XL-10 Gold	Δ (<i>mcrA</i>)183 Δ (<i>mcrCB-hsdSMR</i> - <i>mrr</i>)173 <i>endA1 recA1 relA1 gyrA96 supE44</i> <i>thi-1 lac</i> [F' <i>proAB lacI</i> ^q Z Δ M15 ::Tn10(tet ^R) Amy (Cm ^R)]	High transformation efficiency (Tet ^R and Cm ^R) Expression host	Stratagene
BL21 KRX	F' <i>dcm ompT hsdS_B</i> (r _B ⁻ , m _B ⁻) <i>gal</i> [F', <i>traD36</i> , Δ <i>ompP</i> , <i>proA</i> ⁺ <i>B</i> ⁺ , <i>lacI</i> ^q , Δ (<i>lacZ</i>)M15] Δ <i>ompT</i> , <i>endA1</i> , <i>recA1</i> , <i>gyrA96</i> (Nal ^r), <i>thi-1</i> , <i>hsdR17</i> (r _k ⁻ , m _k ⁺), <i>relA1</i> , <i>supE44</i> , Δ (<i>lac-proAB</i>), Δ (<i>rhaBAD</i>)::T7 RNA polymerase.	Protease deficient High transformation efficiency Expression host	Novagen Promega
<i>Photorhabdus luminescens</i>			
TTO1	Wild Type	Rif ^R	Dr. David Clarke (U.C.C.)
<i>Pseudomonas aeruginosa</i>			
PAO1	Wild Type		Dr Keith Poole

Table 2.2: Plasmids

Plasmid	Description	Source
pCR2.1 (Fig 2.1)	PCR cloning vector, Amp ^R , Km ^R , <i>lacZα</i>	Invitrogen
pUC18	Amp ^R , <i>Plac</i> , <i>lacZα</i> ,	Amersham Pharmacia
pJQ200sk+ (Fig 2.5)	Gm ^R m, <i>mob</i> , <i>sacB</i>	(Quandt and Hynes, 1993)
pBBR1MCS-5 (Fig 2.4)	Gm ^R , <i>mob</i> , Broad host range cloning vector	(Kovach, <i>et al.</i> , 1995)
pUC4K (Fig 6.2)	Amp ^R , Source of Km ^R cassette	Amersham Pharmacia
pRK600	Provides transfer functions, Cm ^R	(Finan, <i>et al.</i> , 1986)
pLecB1	Amp ^R , Expression vector for (His) ₆ tagged LecB, derived from pKK223-3 (Amersham), P _{tac} promoter	Creavin <i>et al</i> , (<i>in prep</i>)
pQE30 (Fig 2.3)	Amp ^R Expression vector for N-terminally tagged (His) ₆ proteins, T5 promoter/ <i>lac</i> operon	Qiagen
pQE60 (Fig 2.2)	Amp ^R Expression vector for C-terminally tagged (His) ₆ proteins, T5 promoter/ <i>lac</i> operon	Qiagen
pCR2.1 derived vectors		
pKM1	pCR2.1 containing <i>lecA</i>	This Study
pKM2	pCR2.1 containing <i>plu2096</i>	This Study
pLecB1 derived vectors		
pLecA1	<i>lecB</i> replaced with <i>lecA</i>	This Study
pLecA4	pLecA1 containing 3 amino acid changes	This Study
pPL-IL	<i>lecB</i> replaced with <i>plu2096</i>	This Study
pQE derived Vectors		
pPL-ILwt	pQE60 containing <i>plu2096</i> with a stop codon encoded before the (His) ₆ tag	This Study
pPL-IL30	pQE30 containing <i>plu2096</i>	This Study
pPL-IL60	pQE60 containing <i>plu2096</i>	This Study
pPA-ILwt	pQE60 containing <i>lecA</i> with a stop codon encoded before the (His) ₆ tag	This Study
pPA-IL30	pQE30 containing <i>lecA</i>	This Study
pPA-IL60	pQE60 containing <i>lecA</i>	This Study
PPA-IL _{mut1} 30	pQE30 containing <i>lecA_{mut1}</i>	This Study
pJQ200sk+ derived vectors		
pKM3	pJQ200sk+ vector containing 5' 1kb region of <i>plu2096</i> region	This Study
pKM4	pKM3 containing 3' 1kb region of <i>plu2096</i>	This Study
pKM5	pKM4 containing a kanamycin cassette from pUC4K	This Study

Table 2.3: Primer Sequences (Synthesised by Sigma-Aldrich, U.K.)

Name	Primer Sequence (5'→ 3')	T _m (°C)
lecA-F1	CCATGGCTTGGAAAGGTGAGGTTCTGG	76.5
lecA-R1	AGATCTGGACTGATCCTTTCCAATATTGACAC	71.3
lecA-F2	GAATTCACGCGTTTTGTGGTGCGCTGGTCATG	83.3
lecA-Rmut1	ACGCGTCGTGGCAGATCAGCCCGTTGTTTCGGAACCTCCCGATCGCCCTG	96.0
lecA-Rwt	ACGCGTCGTGGCAGATCAGCCCGTTGGTCCGGATGCTCCCGATCGCCCTG	97.0
lebBf1	AACGTATGcGGGCCGCAGGGCGAATCGAAA	78.0
lecA _{fntag}	AAAAGGATCCATGGCTTGGAAAGGTGAGG	76.2
lecA _{revntag}	AAAAAAGCTTTTCACGACTGATCCTTTCCAATATT	79.0
Plu2096-F	CTAACCATGGGTTCTGATTGGTCAGGAAGTGT	74.8
Plu2096-R	TTGGAAGATCTTTTTAAAGGGGAGATCGAGACTCT	72.8
PlIIntagfor	ATCCGAGGATCCATGGGTTCTGATTGGTCAGGA	75.1
plIIntagrev	ATCCCAAGCTTTTTAAAGGGGAGATCGAGACTCT	76.2
LecA-Rstop	AAAAAAGATCTTCAGGACTGATCCTTTCCAATATT	70.7
PL-Rstop	AAAAAAGATCTTTATTTTAAAGGGGAGTATCGAGACTC	69.9
lecA6-f	CTCCATTGCGTTTTGTGGTGCGCTGGTCATGAAGATTGGCAACAGCGGAA	92.5
lecA6-r	CTCGGCTGCGGGTTGGTTCGCAACCTCCCGATCGCCCTG	92.1
lecA7-r	CTCGGCTGCGGGTTGGTCCGGATGCTCCCGATCGCCCTG	93.8
lecA8-f	CGTGACAACGTATGGGGGCCGCAGGGCGATCGGGAG	91.5
lecA8-r	TCCGTAACCTGGCCCAACCGGCGGCGACGATGGTAATGACATC	90.2
PL01-f	TCTCAGATAATTCCATGAATTCTCATTACCCT	66.0
PL01-r	CTGCAGCTATTGATGACGGTTGTGGATT	73.4
PL02-f	CTGCAGCCACTCTCGTCGCTAAGATC	73.4
PL02-r	CTCGAGTAAAATTGATATAACCCGGCCAG	70.6
PLc-F	CTCGAGTCGGAGGTAATACTATGTCTGATTG	69.5
PLc-R	TCTAGATTATTTTAAAGGGGAGTATCGAGACTC	66.6
Sequencing Primers		
Q E - F	CCCAAAAGTGCCACCTG	59.6
Q E - R	GTTCTGAGGTCATTACTGG	59.4
K m - f	AATGCAAGTTCTGCATTAGC	59.5
K m - r	TCGATGGTACCAACACAATC	60.5

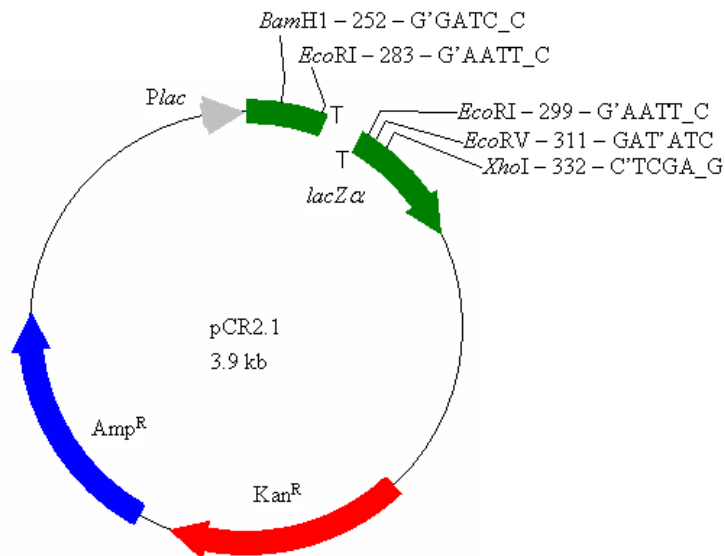


Fig 2.1: pCR2.1 Vector. (Invitrogen, U.S.A.)

The TA cloning vector from invitrogen, with the position of the 3' T overhangs indicated with the letter T. These are located within the *lacZα* ORF, under the control of the P_{lac} promoter (shown in grey). Some of the enzymes located within the multiple cloning site (MCS) are displayed, as well as the position of the ampicillin (blue) and kanamycin (red) antibiotic resistance cassettes.

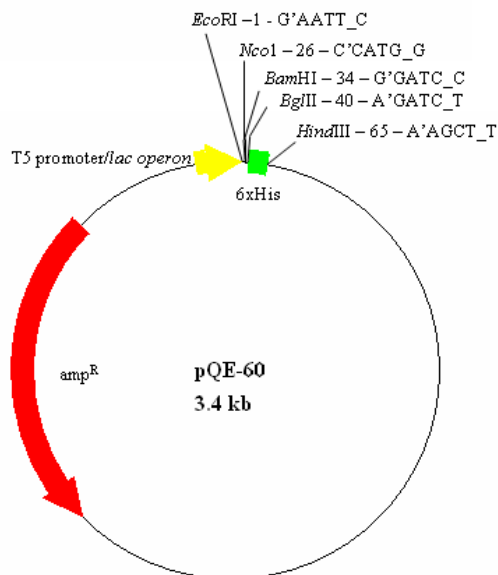


Fig 2.2: pQE60, (Qiagen)

The expression vector pQE60, contains the following features. The MCS is located before the $(His)_6$ amino acid sequence (green) which allows for the C-terminal tagging of a protein with the affinity tag. This is under the control of the T5 promoter/*lac* operon (yellow). The *bla* gene that encodes beta-lactamase confers the bacteria with ampicillin resistance (red).

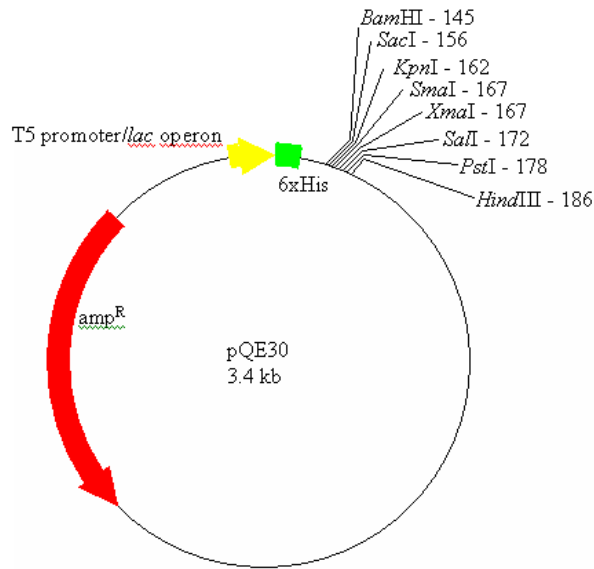


Fig 2.3: pQE30, (Qiagen)

The expression vector pQE30 has the MCS located after the (His)₆ amino acid sequence (green) which allows for the N-terminal tagging of a protein with the affinity tag. This is under the control of the T5 promoter/*lac* operon (yellow). An ampicillin resistance gene also exists on the plasmid (red)

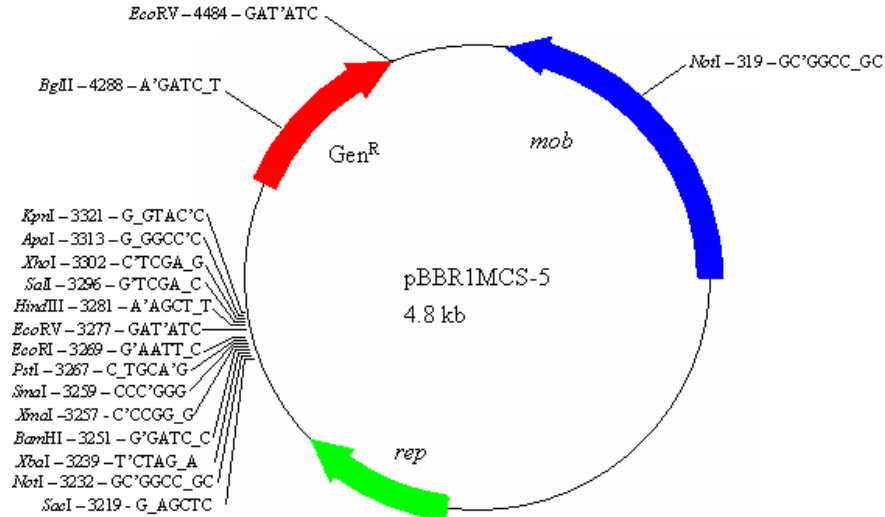


Fig 2.4: pBBR1MCS-5 Vector, (Kovach *et al.*, 1995)

The pBBR1MCS-5 broad host range, has a multiple cloning site (MCS) between bases 3219 and 3321, a gentamycin resistance cassette is shown in red, the origin of replication in green, and the mobilization site in blue.

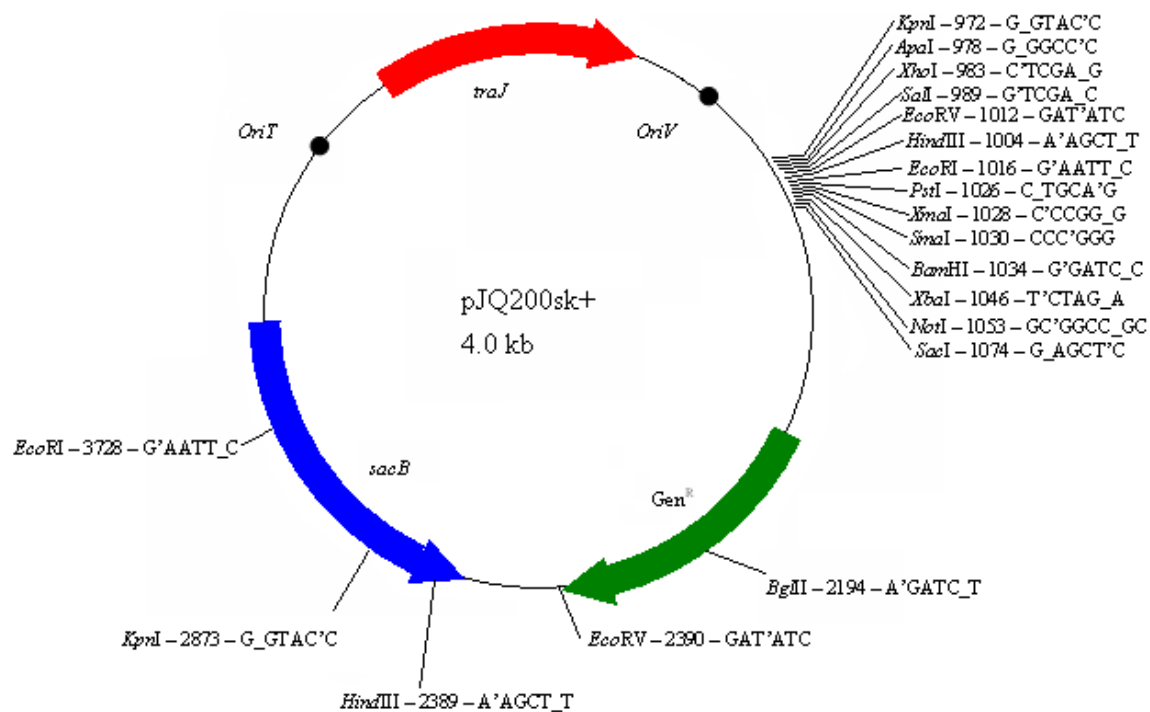


Fig 2.5: pJQ200sk+ vector, (Quandt and Hayes, 1993)

The pJQ200sk+ suicide vector, with the location of the gentamycin resistance gene highlighted in green, and the *sacB* gene shown in blue. The mobilization site (*OriT*) and the origin of replication (*OriV*) are also highlighted. The MCS is located between bases 972 and 1074.

Microbiological media were obtained from Scharlau Ltd unless otherwise stated. Sterilisation was achieved by autoclaving at 121°C and 15 lb/in² for 20 minutes.

Luria Bertani Broth (LB)

Tryptone	10 g/L
NaCl	10 g/L
Yeast Extract	5 g/L

The solution was adjusted to pH 7.0 with NaOH and sterilised by autoclaving. For solid media, 15g/L Oxoid Bacteriological Agar was included. For culturing of *P. luminescens*, sodium pyruvate was added to a final concentration of 0.1% (w/v) (Xu and Hurlbert, 1990).

SOB Medium

Tryptone	20 g/L
Yeast Extract	5 g/L
NaCl	0.5 g/L
KCl	2.5 mM
pH	7.0

The solution was autoclaved and allowed to cool to 55°C before MgCl₂ and MgSO₄ were added to 10mM each from sterile 1 M stock solutions.

2.2 Microbiological media

Nutrient Broth/Agar

Nutrient Broth Powder (Oxoid)	12.8g/L
-------------------------------	---------

Nutrient Agar (Oxoid)	21.4g/L
-----------------------	---------

Sterilised by autoclaving

Lipid Agar

Corn Syrup	10g/L
------------	-------

Yeast Extract	5g/L
---------------	------

Nutrient Agar	25g/L
---------------	-------

Cod Liver Oil	5ml/L
---------------	-------

MgCl ₂ 6H ₂ O	2g/L
-------------------------------------	------

The solution was brought to one litre with dH₂O and autoclaved. Plates were prepared on the same day as inoculation, and poured deeper than normal to ensure moisture retention. In addition, plates were allowed to set with lids in place, and only air dried for a minimal time period.

2.3 Solutions and buffers

TE Buffer

Tris-HCl	10 mM
Na ₂ -EDTA	1 mM
pH	8.0

TAE Buffer (50X)

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

The solution was diluted to 1X with dH₂O before use.

Solutions for 1-2-3 Method of Plasmid Preparation (See section 2.6)

Solution 1

Glucose	50 mM
Na ₂ -EDTA	10 mM (from 0.5 M stock)
Tris-HCl	25 mM (from 1 M stock)

Solution 2:

NaOH	200 mM
Sodium Dodecyl Sulfate	1% (w/v)

Solution 3:

Potassium Acetate	3 M
pH	4.8

To 60 ml of 5 M potassium Acetate, 11.5 ml of glacial acetic acid and 28.5 ml of dH₂O was added. The resulting solution was 3 M with respect to potassium and 5 M with respect to acetate.

TB Buffer

PIPES	10 mM
CaCl ₂	15 mM
KCl	250 mM
pH	6.7

The pH of the solution was adjusted with KOH, and then MnCl₂ was added to 50 mM. The solution was then sterilised through a 0.22 µm sterile filter and stored at 4°C.

Lysis Buffer for Protein Purification

NaH ₂ PO ₄	50 mM
NaCl	300 mM
Imidazole	10 mM
pH	8.0

Solubilisation Buffer for SDS-PAGE

Glycerol	50% (v/v)
SDS	2% (w/v)
2-Mercaptoethanol	5% (v/v)
Bromophenol Blue	0.1% (w/v)
Tris-HCl, pH 6.8	62.5 mM

For native-PAGE, SDS and 2-mercaptoethanol were omitted and replaced with the equivalent volume of H₂O.

SDS-PAGE Running Buffer (5X)

Tris-HCl	125 mM
Glycine	960 mM
SDS	0.5% (w/v)

For native-PAGE, SDS was omitted.

Gel Loading Dye (6X)

Bromophenol Blue	0.25% (w/v)
Xylene Cyanol	0.25% (w/v)
Ficoll (Type 400)	15% (w/v)

Made in dH₂O and sterilised by autoclaving. On a 1% agarose gel, bromophenol blue and xylene cyanol migrate approximately with the 300bp and 4,000bp fragments respectively.

20X SSC Buffer for Colony Blot

NaCl	3 M
Sodium Citrate	300 mM

Adjust to pH 7.0 with HCl and sterilise by autoclaving.

Tris Buffered Saline (TBS) 1X

Tris-HCl	10 mM
NaCl	150 mM
pH	7.4

For TBST, the detergent Triton X-100 was added to a final concentration of 0.1% (v/v).

Phosphate Buffered Saline (PBS) 1X

NaCl	137 mM
NaH ₂ PO ₄	4.3 mM
KCl	2.7 mM
K ₂ HPO ₄	1.5 mM
pH	7.4

For PBST, the detergent Triton X-100 was added to a final concentration of 0.1% (v/v).

Western Blot Transfer Buffer

Methanol	10% (v/v)
Trizma Base	25 mM
Glycine	192 mM

Ethidium Bromide Stain

A 10 mg/ml solution in dH₂O was stored at 4°C in the dark. For the staining of agarose gels, 100µl of the stock solution was mixed with 1 L of dH₂O. The staining solution was kept in a plastic tray and covered to protect against light. Used ethidium bromide was collected and the ethidium bromide was extracted by mixing with a de-staining bag (GeneChoice) overnight. The clear liquid was disposed of routinely, and the ethidium waste was incinerated.

2.4 Antibiotics

- Ampicillin was prepared in dH₂O and stored at -20°C. The working concentration in solid and liquid media for *E. coli* was 100µg/ml.
- Gentamycin was prepared in dH₂O and used at a final concentration of 20µg/ml for *E. coli* and *P. luminescens* in solid and liquid media.
- Kanamycin was prepared in dH₂O and used at a final concentration of 30µg/ml in both solid and liquid media for *E. coli* and *P. luminescens*.
- Tetracycline was prepared in 50% ethanol and stocked at a concentration of 10 mg/ml. This was stored at -20°C. The working concentration for *E. coli* was 10 µg/ml in solid and liquid media.
- Rifampicin was prepared in methanol and a stock solution was prepared at 20 mg/ml. It was stored in the dark at -20°C as it is light sensitive, and used at a concentration of 100µg/ml in solid and liquid media.
- Chloramphenicol was prepared in ethanol at a concentration of 100 mg/ml and stored at -20°C. The working concentration for *E. coli* was 25 µg/ml in solid and liquid media

2.5 Storing and culturing of bacteria

Glycerol stocks were prepared for each strain and stored in duplicate, at -20°C and -80°C. A 0.5 ml aliquot of an exponentially growing culture was added to 0.5 ml of autoclaved 80% glycerol (v/v), and stored accordingly. Where host strains harboured plasmids, the appropriate antibiotic was added to the growth medium. Working *E. coli* and *P. luminescens* stocks were stored on plates at 4°C and room temperature respectively, for a period of up to one week.

2.6 Plasmid preparation by the 1-2-3 Method

This method is adapted from the procedure described by Birnboim and Doly (1979), which utilises the selective alkaline denaturation of high molecular weight chromosomal DNA, while circular plasmid DNA remains double stranded. A 1.5 ml volume of overnight bacterial culture was transferred to a microfuge tube and centrifuged at 13,000 rpm in a microfuge. The supernatant was discarded and the pellet resuspended in 100 µl of Solution 1 (Section 2.3). This was left at room temperature for 5 min. A 200 µl volume of Solution 2 (Section 2.3) was added, mixed by inversion and left on ice for 5 min. To the solution, 200µl of Solution 3 (Section 2.3) was added and the tube mixed by inversion, then placed on ice for 10 min. A clot of chromosomal DNA was pelleted by centrifugation at 13,000 rpm for 10 min. The supernatant was transferred to a fresh tube with 450 µl of phenol chloroform isoamylalcohol (25:24:1) added, which was mixed by rigorous vortexing. Upon centrifugation for 10 min at 13,000 rpm the solution was divided into an upper aqueous layer and a lower organic layer. The upper layer was transferred to a fresh tube and an equal volume of isopropanol was added. After mixing by inversion, the tube was left on ice for 10 min, and then centrifuged for 10 min at 13,000 rpm to pellet the plasmid DNA. The pellet was washed with 100 µl of 70% ethanol and centrifuged at 13,000 rpm for 5 min. The ethanol was removed and the pellet dried in a vacuum dryer for 5 min. The plasmid DNA was resuspended in 50 µl of TE buffer, and stored at 4°C.

2.7 Plasmid preparation using the GenElute plasmid miniprep kit

The kit was used according to the manufacturer's instructions (Sigma-Aldrich). A 1.5 ml volume of an overnight culture was transferred to a microfuge tube and centrifuged at 13,000 rpm for 5 min. The supernatant was removed and the cell pellet re-suspended in 200 µl of re-suspension solution. A total of 200µl of cell lysis solution was then added, and the mixture left at room temperature for 5 min. To this mixture, 350 µl of neutralisation/binding buffer was added and mixed by inversion. The precipitated chromosomal DNA and cell debris was pelleted by

centrifugation for 10 min at 13,000 rpm. A spin column was prepared by spinning 500 µl of spin column preparation solution through it, and discarding the flow-through. The supernatant was then added to the spin column and centrifuged at 13,000 rpm for 1 min to bind the plasmid DNA. The flow through was discarded and the column was then washed by the addition of 750 µl of wash solution, and centrifugation at 13,000 rpm for 1 min. The flow through was discarded and the column dried by centrifugation at 13,000 rpm for 4 min. The column was transferred to a fresh tube and 50 µl of TE buffer added. This was allowed to stand for 5 min before elution of the plasmid DNA by centrifugation at 13,000 rpm for 1 min.

2.8 Preparation of Gram-negative bacterial genomic DNA.

An overnight culture was prepared of the appropriate organism, and 1.5 ml of this culture was pelleted by centrifugation at 13,000 rpm and the supernatant removed. The supernatant was re-suspended in 200 µl lysis buffer, containing 40 mM Tris-Acetate pH 7.8, 20 mM Sodium Acetate, 1 mM EDTA, and 1% SDS (w/v), and the solution mixed by pipetting. To this solution, 66 µl of 5 M NaCl was then added and mixed by inversion. This was then centrifuged at 13,000 rpm for 10 min at 4°C. The solution was then transferred to a fresh tube and an equal volume of phenol chloroform isoamyl-alcohol (25:24:1) added. This was mixed by vortexing and centrifuged as before. The phenol chloroform isoamyl-alcohol purification step was then repeated, and the supernatant removed to a fresh tube containing 2.5 volumes of ice cold ethanol. This was centrifuged as before, and the supernatant removed. The remaining DNA pellet was washed by adding 50 µl of 70% ethanol, and centrifuged as before. The supernatant was removed and the pellet dried in a vacuum dryer. The pellet was then re-suspended in 50 µl of sterile TE buffer, and stored at 4°C.

2.9 Agarose gel electrophoresis

DNA was analysed by electrophoresis through agarose gels in a BioRad horizontal gel apparatus. Agarose was added to TAE buffer to the required concentration (typically 0.7% w/v) and dissolved by boiling. The solution was poured into plastic trays and allowed to set with a plastic comb fitted to create sample wells. TAE was used as running buffer. Loading dye was mixed with the DNA samples to allow loading and to indicate DNA migration during electrophoresis. Gels were run at 140V for ~20 min. Gels were stained by immersion in an ethidium bromide staining solution for 15 min, and then visualised using a UV transilluminator coupled with an image analyser to capture the image on a PC. On every gel 0.5 μ g of 1 Kb Plus DNA Ladder (Invitrogen) was run as a molecular size marker.

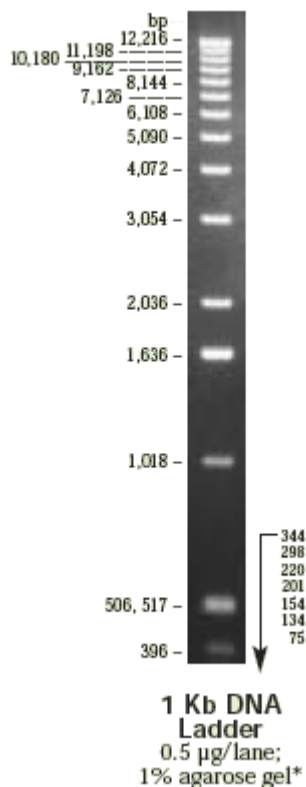


Fig 2.6: 1 kb Molecular Marker (Invitrogen). A representative image of the 1 kb DNA molecular marker used in this study.

2.10 Isolation of DNA from agarose gels

DNA was purified from agarose gels using the DNA Gel Purification Kit (Qiagen), according to the manufacturer's instructions. The samples to be purified were run as normal on an agarose gel that contained 1ng/ml SYBR Safe DNA stain (Invitrogen). The bands were then visualised using a non-UV transilluminator, and the bands excised using a clean scalpel. The gel fragment was placed in a microfuge tube, to which, 500 µl of gel solubilising buffer was added and the tube incubated at 42°C until the gel dissolved. The solution was then transferred to a spin column and was centrifuged at 13,000 rpm for 1 min. The flow through was discarded and 750 µl of wash buffer was then added. The column was then spun as before. The flow through was discarded, and residual ethanol removed by spinning the column at 13,000 rpm for 4 min. The column was transferred to a fresh tube and 30 µl of TE buffer added, and left to incubate at room temperature for 5 min. The column was then centrifuged at 13,000 rpm for 1 min to elute the purified DNA. This was then stored at 4°C until required.

2.11 Preparation of high efficiency competent cells

This method was developed by Inoue *et al* (1990). A glycerol stock of an *E. coli* strain was streaked on LB agar and incubated at 37°C overnight. A 10 ml volume of LB broth was inoculated with a single colony from this plate and cultured overnight at 37°C while shaking at 220 rpm. One ml of this culture was used to inoculate 250 ml of SOB medium in a 1 L flask. This culture was incubated at 37°C, shaking at 220 rpm, until an O.D._{600nm} of 0.6 was reached. The flask was then placed on ice for 10 min. All subsequent steps occurred at 4°C. The culture was transferred to a sterile 250 ml centrifuge tube and centrifuged in a Beckman J2-21 centrifuge at 3,000 rpm for 5 min. The supernatant was discarded and the pellet resuspended in 80 ml of ice-cold TB buffer (Section 2.3). The suspension was left on ice for 10 min and centrifuged as before. The resulting cell pellet was finally resuspended in 20 ml of ice-cold TB buffer and DMSO was added to a final

concentration of 7%. After incubation on ice for 10 min the cells were aliquotted into cooled sterile microfuge tubes and flash frozen in liquid ethanol before storage at -80°C.

2.12 Transformation of competent cells

An aliquot of the competent cells was thawed on ice. A 200 µl aliquot of the cells was mixed gently with 2µl plasmid DNA/ligation mixture and kept on ice for 30 min. The cells were heat shocked at 42°C for 30 seconds and quickly placed back on ice for 2 min. Using aseptic technique, 800µl of LB broth was then added to the cells followed by incubation at 37°C for 60 min. A total of 100µl of the transformation suspension was then spread on an LB agar plate containing the relevant antibiotics and incubated at 37°C overnight.

2.13 Determination of competent cell efficiency

Competent cell efficiency was defined in terms of the number of colony forming units obtained per µg of transformed plasmid DNA. A 25 ng/µl stock of pUC18 plasmid DNA was diluted to 250 pg/µl, 25 pg/µl, and 2.5 pg/µl. An aliquot of 2 µl from each dilution was transformed as described in Section 2.12. The transformation efficiency was calculated from the number of colonies obtained, taking into account the dilution factor and the volume of each culture transferred to the spread plate.

2.14 Protein expression

An LB plate with the appropriate antibiotic was streaked with a glycerol stock of the strain containing the expression plasmid. One colony was selected and used to inoculate 5 ml of LB broth containing the appropriate antibiotic, and grown overnight at 37°C while shaking. A 250 ml conical flask containing 100 ml of LB broth was inoculated with 1ml of the overnight culture, and the antibiotic added. The culture was incubated at 37°C, with shaking at 220 rpm until an OD (A_{600}) of 0.6 was reached. IPTG was then added to a concentration of 50 μ M to induce expression. The culture temperature was then dropped to 30°C and incubated for a further 6 hours, whereupon the cells were harvested. The culture was centrifuged at 5,000 rpm for 5 min (using a Sorvall SLA-1500 rotor) to pellet the cells. The supernatant was discarded and the cell pellets then stored at -20°C until lysis.

2.14.1 Cell lysate preparation

Cell pellets from Section 2.14 were washed with 50 ml of lysis buffer (Section 2.3). This was centrifuged at 5,000 rpm for 5 min and the supernatant discarded. The cells were then re-suspended in 10 ml lysis buffer, and disrupted using a 3 mm micro-tip sonicator (Sonics & Materials Inc). Cells were subjected to 2.5 sec, 40 kHz pulses for 30 sec while being kept on ice. DNase was added to a concentration of 10 μ g/ml, and cell debris removed by centrifugation at 4,000 rpm for 20 min at 4°C. The cleared lysate was filtered through a 0.45 μ m sterile filter into a sterile universal and stored at 4°C.

For large scale purifications, the cell pellet pre-sonication was re-suspended in 10 ml lysis buffer per 100 ml culture. The re-suspended pellet was then treated with 10 mg/ml lysozyme at 37°C while shaking gently for 30 min before sonication on ice. DNase treatment and filtration was as normal, with sonication time increased to 30 sec per 100 ml culture.

2.14.2 Colony blot procedure

Using a multichannel pipette, 500 µl of LB containing the required antibiotic was aseptically aliquotted into an autoclaved deep well 96-well plate. Each well was inoculated with a single colony obtained from the transformation plate. The plate was sealed with Breathe-Easy Sealing Membrane (Sigma), and incubated at 37°C overnight. An LB-amp agar plate containing 50 µM IPTG was overlaid with nitrocellulose membrane, and an equal volume each of overnight culture spotted on the membrane. This was incubated at 37°C for eight hours. The nitrocellulose membrane was then removed carefully and placed colony side up on blotting paper that had been soaked in 10% (w/v) SDS, with care taken not to introduce any air bubbles. After 5 minutes the nitrocellulose membrane was removed and placed colony side up on blotting paper that had been soaked in denaturing solution (0.5 M NaOH, 1.5 M NaCl). After 5 min incubation the membrane was again removed and placed on blotting which had been pre-soaked in neutralization solution (1.5 M NaCl, 0.5 M Tris-Cl, pH 7.4). This step was repeated, and finally the membrane was placed on a solution containing 2X SSC solution (Section 2.3).

The membrane was washed twice in TBS (Section 2.3), and blocked for one hour in TBS containing 2.5% BSA. Unbound BSA was removed by washing twice in TBST (Section 2.3), before the addition of 1:10,000 mouse anti histidine HRP-labelled antibody. This was allowed to bind for one hour before washing unbound antibody twice for 10 minutes in TBST. Two Sigma colourfast tablets were dissolved in 100ml of dH₂O, and poured on the membrane. This was left to develop for 2-5 min, and stopped by rinsing twice with dH₂O.

2.15: Immobilised metal affinity chromatography (IMAC)

2.15.1 IMAC using Ni-NTA resin

A 1 ml aliquot of nickel-nitrilotriacetic acid resin (Ni-NTA, Invitrogen) was mixed with 10 ml of cleared cell lysate (Section 2.14) at 4°C overnight. The mixture was then poured into a 0.7 x 15 cm column allowing the resin to settle. The column was then washed with 10 ml of cell lysis buffer (Section 2.4). Two 20 ml washes of lysis buffer supplemented with up to 80 mM imidazole was then passed through the column, with elution of the final protein by the addition of 5 ml lysis buffer supplemented with 300 mM imidazole. Imidazole was then removed by the transfer of the eluent to an Amicon Centrifuge tube. The solution was de-salted in the Amicon tube by centrifugation at 3,800 rpm for 20 min, and the protein was washed twice in PBS (Section 2.3), and concentrated to a final volume of 1 ml. Sodium azide was then added to a concentration of 0.05% (w/v), and frozen, or the sample was lyophilized if long-term storage was necessary.

2.15.2 IMAC using FPLC and Amersham nickel-resin

Before the attachment of the HisTrap FastFlow Crude 1 ml column to the FPLC the maximum back pressure was set to 3 bar. The top of the affinity column was attached to the pump outlet no 3 and the bottom of the column attached directly to the FPLC unit. The HisTrap column storage buffer (20% ethanol) was washed out with 5 column volumes of water (5 ml) and the resin equilibrated with at least 10 column volumes of lysis buffer (Section 2.3) at a flow rate of typically 1 ml/min.

A blank run was then performed using the same purification protocol that was required for the sample. The sample injection loop was filled with a volume of lysis buffer that was equal to the sample that was purified. A stock of lysis buffer was attached to pump A, with lysis buffer containing 1 M imidazole attached to pump B. The purification strategy involved

- Column equilibration with 10 ml Lysis buffer
- Sample application from sample injection loop
- A wash of 20 ml lysis buffer
- Application of an imidazole gradient to remove non-specifically bound proteins, by applying 20 ml of a 20 mM imidazole wash
- A more stringent wash of 80 mM imidazole was then applied (20ml)
- Purified protein was then eluted by the application of 1 M imidazole

After the blank run, the sample was applied, and the process repeated. The absorbance values measured during the blank run were subtracted from the absorbances recorded in the sample run, and the sample values were determined.

2.15.3: Desalting of purified protein using HiPrep 26/10 desalting column

The FPLC maximum post column pressure limit was set to 3.5 bar before attachment of the HiPrep 26/10 Desalting Column. The top of the column was attached to pump outlet no 5 and the bottom of the column attached to pump inlet no. 2. Before the first sample application the column storage buffer (20% ethanol) was removed, and the column equilibrated with sample buffer. This was done by washing with 2 column volumes of water (106 ml) followed by 5 column volumes (265 ml) of sample buffer (either water or PBS) (Section 2.3). Equilibration was not necessary between runs using the same buffer. Washing was done at the flow rate intended for chromatography, typically 20 ml/min. Samples were collected in 1 ml fractions in 96-well plates.

2.16 Recharging of IMAC resin

The column was routinely recharged prior to re-use. The used resin was poured into a column and washed with 2 column volumes of dH₂O, followed by 2 column volumes of 50% (v/v) ethanol. The resin was then stripped with 2 column volumes of 100 mM EDTA, pH 8.0. Remaining impurities were removed with 2 column volumes of 200 mM NaCl, followed by 2 column volumes of dH₂O. Hydrophobically bound proteins and lipoproteins were removed by washing with 10 column volumes of 30% isopropanol for 30 minutes, followed by 10 column volumes with water. The resin was then recharged by adding 2 column volumes of 100 mM NiSO₄. The resin was washed again with 2 column volumes of dH₂O, before transfer to a sterile universal, where it was stored in 20% ethanol.

2.17 Protein quantification by BCA assay

The bicinchonic acid (BCA) described by Smith *et al*, (1985) was utilised to quantify total protein in the range of 20-2,000 µg/ml. All samples were diluted to within the range of the assay and added in triplicate to 150 µl of BCA reagent (Sigma), and incubated at 37°C for two hours. Absorbance was read at 562 nm. A standard curve was created using bovine serum albumin (BSA) as the reference protein (Appendix A). Protein concentration of the unknown was determined from this standard curve.

2.18 SDS-PAGE

The SDS-PAGE method was used to analyse protein samples. A 20% resolving gel and a 4% stacking gel were prepared as per Table 2.4. Gels were cast using the ATTO vertical mini electrophoresis system. Upon the addition of TEMED the resolving gel was immediately overlaid with isopropanol. After polymerisation, the isopropanol was poured off and the stacking gel added. A comb was inserted into the top of the gel to form loading wells.

Table 2.4: SDS-PAGE gel recipes

Solution	20% Resolving Gel	4% Stacking Gel
1.5 M Tris-HCl, pH 8.8	1.875 ml	-
0.5 M Tris-HCl, pH 6.8	-	0.94 ml
dH ₂ O	0.54 ml	2.3 ml
Acrylamide/Bis-acrylamide 30%/0.8% (w/v)	5.0 ml	0.5 ml
10% (w/v) Ammonium Persulphate	38 µl	19 µl
20% (w/v) SDS	38 µl	19µl
TEMED	4 µl	2 µl

2.18.1 Sample preparation:

To a microfuge tube, 20µl of sample was added to 5 µl of 5X solubilisation buffer (Section 2.3). Samples were then boiled for 5 min and applied to wells that were flushed of un-polymerised acrylamide. The preparation of insoluble fractions depended on the original sample. The insoluble fraction should be applied in the same concentration as the soluble fraction, so to the insoluble pellet obtained from a 1 ml sample, 20µl of 5X solubilisation buffer was added, and boiled for 5 min. The sample was then diluted with water to the original 1 ml with dH₂O.

2.18.2 Sample application

A total of 25µl of the prepared sample was applied to each SDS-PAGE well. One lane on each gel was kept for the relative molecular weight protein marker (SigmaMarker, Sigma, Fig 2.7). A 10µl aliquot of the marker was applied, which consisted of Rabbit Muscle Myosin (205 kDa), β-Galactosidase from *E. coli* (116 kDa), Rabbit Muscle Phosphorylase b (97 kDa), Rabbit Muscle Fructose-6-phosphate Kinase (84 kDa), Bovine Serum Albumin (66 kDa), Bovine Liver Glutamic Dehydrogenase (55 kDa), Chicken Egg Ovalbumin (45 kDa), Rabbit Muscle Glyceraldehyde-3-phosphate Dehydrogenase (36 kDa), Bovine Erythrocyte Carbonic Anyhdrase (29 kDa), Bovine Pancreas Trypsinogen (24 kDa), Soybean Trypsin Inhibitor (20 kDa), Bovine Milk α-Lactalbumin (14.2 kDa) and Bovine Lung Aprotinin (6.5 kDa). When western-blot analysis was

required, the NEB pre-stained protein ladder was used. A representative image of both ladders is given in Fig 2.5. Gels were run at 25 mA for 70 min at room temp with 1X SDS-PAGE running buffer (Section 2.3).

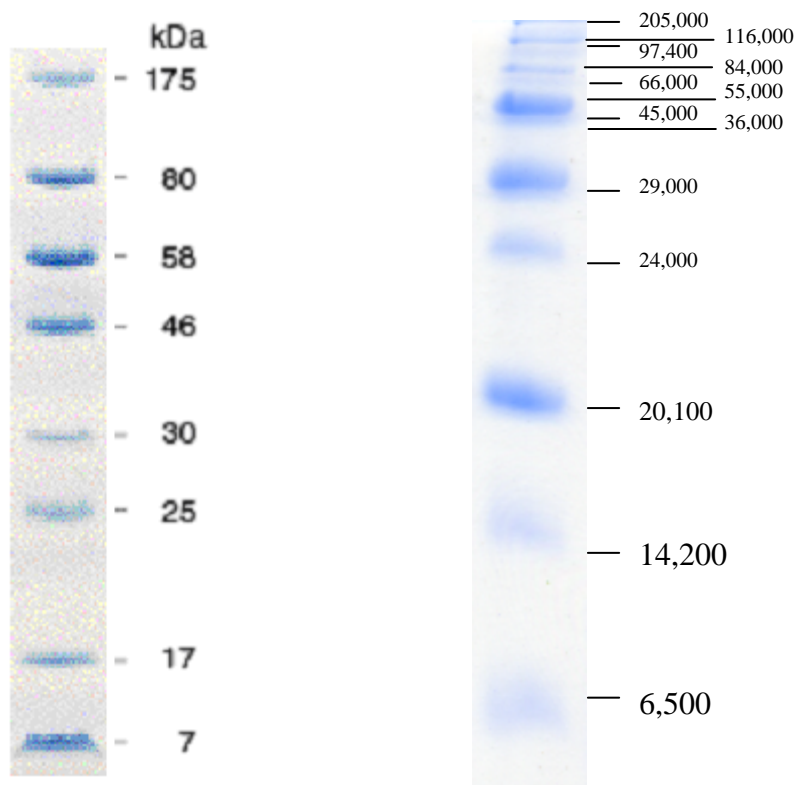


Fig 2.7. NEB Prestrained marker and Wide Range SigmaMarker visualised on 20% SDS-PAGE.

2.18.3 Gel analysis

PAGE gels were removed from the electrophoresis chamber and washed with dH₂O. Gels were then routinely stained for one hour with staining solution which contained 10% (v/v) Acetic Acid, 45% (v/v) Methanol and 0.25% Coomassie Blue. Overnight destaining followed using destaining solution which consisted of 10% (v/v) Acetic Acid and 45% (v/v) Methanol. For gels which required a greater degree of sensitivity the silver staining method was used which was described by Blum *et al* (1987), and outlined in table 2.5.

Table 2.5: Silver Staining of SDS-PAGE gels

Step	Duration	Reagent (total volume for each, 100 ml)
Fix	60 min (min)	30 % Ethanol, 10 % Acetic Acid
Wash	15 min	20 % Ethanol
Wash	15 min	dH ₂ O
Sensitize	1 min	0.1% (w/v) Na ₂ S ₂ O ₃ ,
Rinse	2 x 20 sec	dH ₂ O
Silver Stain	20 min	0.1 % AgNO ₃ , 70 µl of 37% (v/v) Formaldehyde stock
Rinse	2 x 20 sec	dH ₂ O
Develop	10 min (max)	3% Na ₂ CO ₃ , 50 µl of 37% (v/v) Formaldehyde stock, 0.002% (w/v) Na ₂ S ₂ O ₃
Stop	5 min	50 g/L Trizma Base. 2.5 % Acetic Acid (v/v)

2.19 Native-PAGE

For the preparation of Native-PAGE gels, the concentrations outlined in Table 2.4 were used with the omission of 20% SDS, and supplemented with the appropriate volume of water. The sample to be analysed was not boiled and the sample buffer contained neither SDS nor 2-Mercaptoethanol. The gels were run in 1X Native-PAGE running buffer (Section 2.3).

2.20 Western blot

An SDS-PAGE gel was run (Section 2.18.2) using the NEB pre-stained molecular weight marker. Four pieces of blotting paper, and a piece of nitrocellulose membrane were cut to the same dimensions as the SDS-PAGE gel. These were soaked in transfer buffer (TBS), and arranged on the semi-dry electroblotter as per Fig 2.8. Transfer then occurred at a constant 50 mA for 1 hour.

To detect transferred recombinant (His)₆ tagged proteins, the membrane was washed twice in TBST, and then blocked with 2.5% BSA-TBST (Section 2.3) for one hour. The membrane was washed twice more in TBST and then incubated with 1:10,000 murine anti-His antibody in 1% BSA-TBST. This was finally washed four times with TBST and incubated in 15 ml sH₂O containing SIGMAFAST™ 3,3'-Diaminobenzidine tablets. Development was stopped by washing with sH₂O.

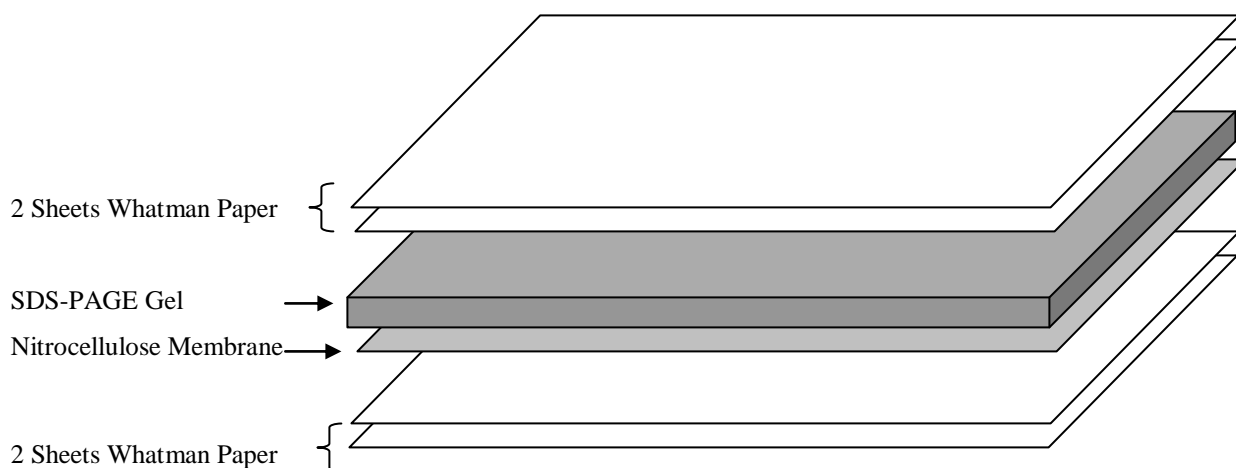


Fig 2.8: Schematic of Western Blot. Sheets of Whatman paper and nitrocellulose membrane are cut to the exact same dimensions as the SDS-PAGE gel, as any overlap over the gel surface can lead to inefficient protein transfer.

2.21 Gel filtration

Before the attachment of the Superdex 75 10/300 High Performance Column (GE Healthcare; Uppsala, Sweden) to the AKTA Purifier FPLC (Amersham Biosciences; Uppsala, Sweden), the maximum back pressure limit was set to 1.8 mPa. The column storage valve was disconnected from the gel filtration column, and the top connected to pump outlet no. 1. The bottom of the column screwed directly into the FPLC. To remove the column storage buffer (20% ethanol) two column volumes of water were pumped through the column (Pump P-900, Amersham Biosciences; Uppsala, Sweden) at a flowrate of 1 mL/min. The column was then equilibrated with 5 column volumes of sample buffer at the same flow rate, before sample application. With pump A connected to a stock of degassed sample buffer, 1 mg of sample was applied through the sample injection port in a volume of no more than 100 μ l, and the run commenced, with a constant flow rate of 1 ml/min of buffer A. After sample addition, 2.5 column volumes of buffer was passed through the column. The proteins retention time was measured using the online Monitor UV-900 (GE Healthcare; Uppsala, Sweden), which read the eluent absorbance at OD 280nm.

2.22 Blood preparation for hemagglutination assay

Whole blood was obtained from a rat and stored in 1% (v/v) sodium citrate, for a maximum of 3 days. To wash the red blood cells (RBCs) 500 μ l of blood was mixed with 500 μ l PBS, and then centrifuged at 600 g for 10 min at 4°C. The supernatant was removed and the remaining red blood cells re-suspended in 1 ml of PBS. The washing process was repeated three times. To remove glycoproteins from the surface of the RBCs, 350 μ l of RBCs was mixed with 9.65 ml PBS containing 0.1% papain and 0.01% cysteine. This solution was incubated for 30 min at 37°C with shaking. The RBCs were then spun down as before, and washed four times with PBS. The cells were finally resuspended in 965 μ l PBS, giving a 3.5% RBC solution. This was stored in at 4°C for a maximum of 5 days.

2.23 Hemagglutination assay

A 25 µl aliquot of PBS was added to each well of a U-bottomed 96-well plate. Purified lectin in a total volume of 25 µl was then added to one well containing 25 µl PBS, which was mixed, and 25 µl of the solution was transferred to the next well creating a serial dilution across the plate. A total of 50 µl of 3.5% papain-treated RBCs was then added to each well, along with a further 25 µl of PBS, and was left to incubate for 30 min at room temperature. Hemagglutination was observed as a coating of RBCs across the bottom of the well, whereas no agglutination activity was observed as a spot of RBCs which had aggregated as a result of gravity. One hemagglutination unit (HU) was defined as the minimum quantity of lectin required to agglutinate the RBCs.

2.23.1 Hemagglutination inhibition assay

A serial dilution of the relevant sugar was created on a U-bottomed 96-well plate using a stock sugar solution in PBS. A total of 8 HU of the lectin in 25 µl was then added, and allowed to incubate for 20 min. A 50 µl aliquot of RBCs were then added to each well, and incubated for 30 min at room temperature.

2.24 Enzyme linked lectin assay

A 50 µl volume of glycoprotein was immobilised in each well of a 96-well Nunc MaxiSorb ELISA plate. Each sample was assayed in triplicate, at a concentration of 10 µg/ml. These were left overnight at 4°C, or if glycoprotein denaturation was required, the sample was heated to 60°C for 30 min, and then applied to the plate. The plate was then immobilised at 60°C overnight. The unbound glycoprotein was removed by inverting the plate and the wells were then blocked with 150 µl of 5% BSA in TBS for one hour. This solution was then removed by inverting the plate

and was washed using TBS supplemented with 0.1% Triton X. A 50 μ l aliquot of lectin suspended in TBST-BSA 1%, supplemented with 10 mM CaCl_2 was then added at a concentration of 10 $\mu\text{g/ml}$, and let incubate at room temperature for one hour. This was removed by inversion and washed with TBST. This was followed with 50 μ l of 1:10,000 mouse anti-histidine or anti-biotin antibody as appropriate. This was created fresh and diluted with TBST-BSA 1%, and incubated for one hour at room temperature. Unbound antibody was removed by inversion and washed four times with TBST, before the addition of 50 μ l of TMB substrate (Section 2.3). The reaction was stopped after 5 min by adding dilute H_2SO_4 and the absorbance was read at 450nm.

2.25 Lectin purification using sepharose-4B

Sepaharose-4B was poured into a 20 ml column and washed with 2 column volumes of water to remove the storage buffer. The column was then equilibrated by washing with two column volumes (C.Vs) of lysis buffer. Cleared cell lysate prepared as outlined in Section 2.14.1, was then added to the column, which was allowed to mix overnight at 4°C. Unbound material was collected as flowthrough, followed by a 20 ml wash using lysis buffer. Lectin elution was then facilitated by adding 20 ml of EDTA 0.2 M pH 8.0.

2.26 Immobilisation of protein onto cyanogen bromide-activated agarose

The protein to be immobilised was suspended in 0.1 M NaHCO₃ coupling buffer, pH 8.5, which contained 500 mM NaCl. The capacity for the gel was 5-10 mg per ml of gel. The resin was then washed in cold 1 mM HCl for 30 mins, at 200 ml per g of gel. The supernatant was removed and the resin was washed with 10 cv of dH₂O, followed by 1 cv of coupling buffer. The protein solution was immediately added to the column, and the flow stopped. The solution was allowed to mix by inversion overnight at 4°C, after-which it was washed with coupling buffer to remove unbound protein. The resin was then blocked by adding 2 cv of 0.2 M glycine, pH 8.0, and was allowed to mix by inversion for 2 hours at room temperature. The unbound glycine was removed by washing with 2 cv coupling buffer, followed by a wash of 2 cvs using 0.1 M Sodium Acetate, pH 4.0, containing 500 mM NaCl. Alternate washing steps using coupling buffer and acetate buffer were repeated 5 times. The resin was then used immediately or stored at 4°C in 1 M NaCl with 0.05% sodium azide.

2.27 Tri-parental mating

P. luminescens was grown to late stationary phase in nutrient broth, while two *E. coli* cultures containing the donor plasmid and the mobilising plasmid pRK600 (Finan *et al.*, 1986) were also grown to stationary phase. The two *E. coli* cultures (0.75 ml each) were mixed and pelleted by centrifugation at 13,000 rpm for 3 min, and resuspended in 100 µl of fresh LB. This suspension was then spotted onto the centre of an LB plate. This was incubated overnight at 37°C and then resuspended in 3 ml of LB broth. A 0.75 ml aliquot of this culture was then pelleted as before, with 0.75 ml of *P. luminescens* recipient culture. These were re-suspended in 100 µl of nutrient broth and spotted onto the centre of a nutrient agar plate. This was left incubate at 30°C for two nights, before re-suspension in 2 ml of nutrient broth. Dilutions of this suspension were then spread on selective media, along with aliquots from each previous step, to act as controls.

2.28 In silico analysis of DNA and protein sequences

To identify homologous protein and DNA sequences deposited in GenBank, the BLAST programme (Altschul *et al.*, 1997) available at NCBI (www.ncbi.nlm.nih.gov) was used. DNA sequences for the strains used in this study were obtained from the following sequencing studies; *Pseudomonas aeruginosa* PAO1 (NC_002516), *Photobacterium luminescens* (NC_005126). DNA and protein sequences were aligned using the ClustalW programme (Thompson, 1994) available at (<http://www.ebi.ac.uk/Tools/clustalw/>) and the Genedoc programme, available to download at <http://www.nrbsc.org/>. DNA sequences were analysed for restriction sites using the Webcutter 2.0 programme (<http://rna.lundberg.gu.se/cutter2/>). Protein imaging software used was Deepview Swiss PdbViewer from <http://www.expasy.org/spdbv/> and PyMol from Delano Scientific available from <http://pymol.sourceforge.net/>.

2.29 TA cloning of PCR products

Polymerase chain reactions (PCR) carried out with RedTaq (Invitrogen) DNA polymerase were routinely performed to clone a gene of interest into the pCR2.1 vector. Fig 2.9 shows the principle of the TA cloning method.

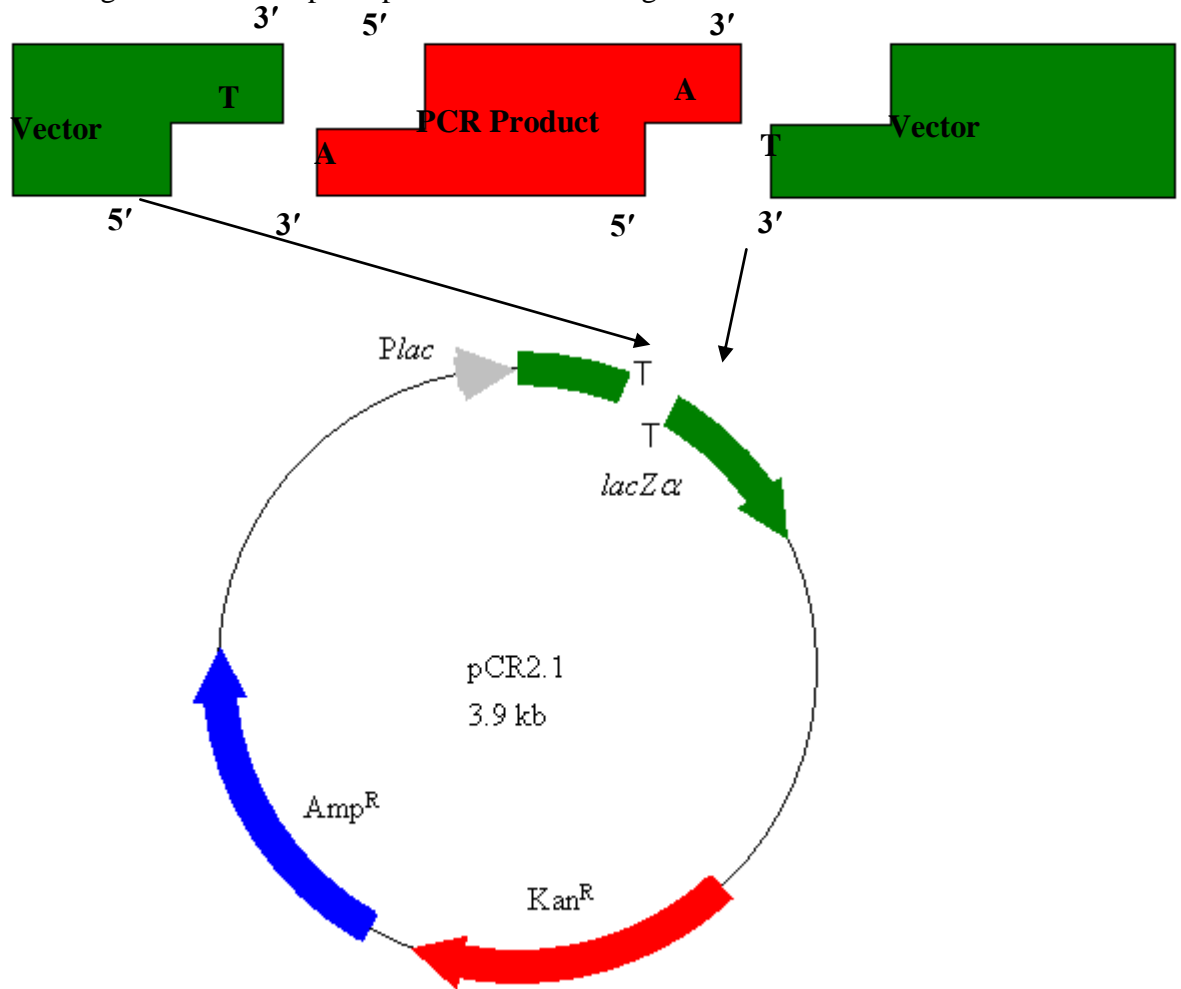


Fig 2.9: Principle of TA Cloning

PCR products amplified by RedTaq polymerase have a deoxyriboadenosine addition at the 3' end. The commercial pCR2.1 vector has been linearised within the *lacZα* gene, and the cleavage site contains deoxyribothymidine overhangs. This is complementary to the dA of the PCR product, and can thus be ligated easily.

2.30 Enzymatic reactions:

Standard Phusion PCR Reaction Mixture:

Template DNA	1 μ l
Primers	0.5 μ l of each
Buffer (5X)	4 μ l
dNTPs (10 mM each)	4 μ l
dH ₂ O	9 μ l
PhusionTaq Polymerase (1:5 dilution)	1 μ l

Standard PCR Program Cycle

Stage 1:	Step 1: 98°C for 5 min
Stage 2: (30 Cycles)	Step 1: 98°C for 10 sec Step 2: Annealing temp for 15 sec (T _{ann} was routinely 5 below the T _m) Step 3: 72°C for 20 sec for every kb to be synthesised
Stage 3:	Step 1: 72°C for 10 min

RedTaq PCR Reaction Mixture

Template DNA	1 μ l
Primers (0.5 nm/ μ l)	1 μ l of each
RedTaq Buffer (10X)	5 μ l
dNTP mix (10 mM)	1 μ l
Sterile dH ₂ O	40 μ l
RedTaq DNA Polymerase	1 μ l

RedTaq (Invitrogen) PCR Program Cycle

Stage 1:	Step 1: 95°C for 10 min
Stage 2: (30 cycle)	Step 1: 95°C for 1 min
	Step 2: Annealing temp for 30 sec
	Step 3: 72°C for 1 min per kb to be synthesised
Stage 3:	Step 1: 72°C for 10 min

TA Ligation Reaction:

PCR Product	1 µl
pCR2.1 Vector	1 µl
T4 DNA Ligase	1 µl
10X Ligase Buffer	1 µl
dH ₂ O	6 µl

2.31 ElectroSpray ionization mass spectrometry

Fresh samples (< 3 days old) of lectins were used as lectins tend to aggregate in solution over time. Samples were desalted using Centricon PM10 (mass cut off at 10,000 Da) micro-concentrators (Amicon, Millipore, Bedford MA, USA) and diluted to appropriate concentrations with mass spectrometry grade solvents.

Electrospray ionization mass spectrometry (ESI-MS) was performed on a quadrupole- time-of-flight mass spectrometer (Q-TOF Ultima Global™, Micromass, Manchester, UK) coupled with either a standard or nanoflow Z-spray source. For the standard source, desolvation temperature was set at 65°C. The nanoflow source interface was operated at room temperature and was equipped with a PicoTip™ emitter with distal coating (20 µm i.d with 10 µm Silica tip™, New Objective, Woburn, MA, USA). Ions were focused by a radio frequency (RF) lens before transmission to the quadrupole, which was used in RF-only mode as a wide bandpass filter. The ions were then transmitted through the hexapole collision cell and were

pulsed into the TOF analyzer. The TOF analyzer was set in the positive ion mode at 9.10 kV when the reflectron was used in V configuration at 35.6 V and pusher was in auto mode. In W configuration, the TOF analyzer was set at 10.15 kV with the reflectron at 25 V and the pusher was set manually at 250 μ s. Acquisition was achieved by using a time-to-digital converter operating at 4 GHz. Mass spectra were recorded using a micro channel plate (MCP) detector at the exit of the TOF analyzer, after calibration with NaI solution (1 mg/ml) in isopropanol/water (50/50). A mass range (m/z) of 500 – 5000 Th was scanned over 5 sec with an inter scan delay time of 0.1 sec. Samples were analyzed in both “native like” and “denaturing” conditions and instrument parameters were optimized separately for each condition as described in the results and discussion section. Samples (typically 5 - 10 pmol/ μ l) were introduced into the source by direct infusion using a syringe pump (Cole-Parmer, Vernon Hills, Illinois, USA) at a rate of 0.8 ml/min for the standard source and 200 nl/min for the nanoflow source.

Calculations and data analyses were performed using MassLynx software version 4.0 (Waters, Manchester, UK). Deconvolution of spectra was accomplished either manually or by using a transform algorithm. Mass spectra were background subtracted, smoothed using the Savitzky Golay method, and centered before final calculations of the molecular mass. Molecular species were represented by an envelope of a series of peaks corresponding to multiply charged ions that deconvoluted to a mean mass \pm standard deviation.

2.32 *P. luminescens* pathogenicity assay

One ml of *P. luminescens* overnight culture was centrifuged at 8,000 rpm for two minutes. The supernatant was discarded and the pellet washed in 1 ml of PBS (Section 2.3) and centrifuged as before. The final pellet was re-suspended in one ml of PBS, and the OD measured at 600nm. The cells were diluted in PBS and the OD measured at 600nm. The cells were then diluted in PBS to give a final O.D. of 1.0 which equates to 2×10^8 cells/ml. This was then diluted to 1×10^4 cells/ml. A hypodermic syringe was sterilised by submersion in 100% ethanol, and flushed several times. The syringe was then flushed with sterile PBS 8 times. Ten *Galleria*

mellonella larvae were injected between the 1st and 2nd set of prolegs, with either sterile 10µl of PBS solution to act as negative controls, or by 10µl of the 10⁴ cells/ml dilution. The insects were placed in a 25°C incubator and monitored for the next 48 h, checking every 1.5 h for insect death from the 42 hour time-point.

For a competitive assay, two strains were diluted to an OD_{600nm} of 1.0, and then mixed in a ratio of 50:50. This mix was then diluted to 1x10⁴ cells/ml and injected into the insect as per the standard pathogenicity assay. The dilutions were plated on plates containing the appropriate antibiotics to ensure the ratios were correct.

To check the ratios of bacteria within the dead insects, they were surface sterilised by dipping in ethanol and excess liquid was removed with blotting paper. The cadaver was flamed quickly, before being placed in a sterile Petri dish containing 5 ml of PBS. A sterile scalpel blade was then used to cut the insect lengthwise into two, and the resulting solution was poured into a universal containing 7 sterile glass beads. The mixture was mixed in a vortex for two minutes, and the resulting solution serially diluted in PBS. This was spread on plates containing the appropriate antibiotics to ascertain relative numbers and hence their competitiveness.

2.33 *P. luminescens* symbiosis assay

A total of 50µl of *P. luminescens* overnight culture was diluted to an O.D_{600nm} of 1.0 and spotted three times onto freshly poured lipid agar plates (Section 2.3). Using a sterile loop, this was spread across the plate in a 'Z' pattern. These were inverted and incubated at 30°C for three days. A flask of TT01 *Heterorhabditid* nematodes, at least two-weeks old, were checked for viability under the microscope. Live nematodes are bent and motile, dead nematodes are straight. Approximately 20 ml of nematodes were poured into a sterile 50 ml universal, and PBS was added to a final volume of 45 ml. A 5 ml aliquot of 4% hyamine was added to the suspension, and left for 15 minutes, after which as much as possible of the supernatant was removed. Then 50 ml of PBS was added, and the nematodes were allowed to settle for 15 min. The supernatant was decanted and PBS added again to 50 mls. This wash step was repeated three times before finally re-suspending the surface-sterilized nematodes in

50 ml of PBS. To prevent damaging the nematodes, the point of a P20 tip was removed, and 10 μ l of the solution dropped onto a petri dish three times. The number of nematodes was counted under a microscope to calculate an average number of nematodes per 10 μ l. A volume equivalent to 50 nematodes was then added to the centre of the symbiosis plate. The plate was wrapped in parafilm, and stored at 25°C for the remainder of their life-cycle.

For a competitive symbiosis assay, two strains were mixed 50:50 once their OD_{600nm} had been adjusted to 1.0. Control plates were required for each ratio set up. To achieve this the cultures collected at 3 and 21 days were plated on media containing the relevant antibiotic to ensure the mixing ratio was correct.

After 21 days, live infective juveniles (IJs) had migrated to the lid of the plate, from where they were washed off in PBS, and counted. They were surface sterilised in hyamine as before, after which one individual nematode was selected and added to a 1.5 ml centrifugation tube. The nematode was re-suspended in 100 μ l of PBS, and crushed by micro-pestle for one minute. Then 50 μ l of the resulting extract was spread on plates containing the appropriate antibiotic.

3.0 Cloning Expression and Purification of Recombinant PA-IL and PL-IL

3.1 Introduction

This chapter describes the cloning of the two genes that encode the lectins PA-IL and PL-IL in the organisms *P. aeruginosa* and *P. luminescens*, and the subsequent development of recombinant expression systems, which were then utilised to prepare active recombinant protein to a level of high purity. A mutant derivative of the PA-IL lectin was also constructed. Structural (Chapter 4) and functional analysis (Chapter 5) of these lectins was then carried out.

3.2 Cloning, small-scale expression and mutagenesis of lectin genes

3.2.1 Cloning and small scale expression of the *P. aeruginosa* gene *lecA* encoding PA-IL

The *P. aeruginosa* strain PAO1 was obtained from Prof. Keith Poole, and the sequence for the organism (Fig 3.1) obtained from the NCBI data bank (data entry no NC002516). The primers used for the amplification of *lecA* were lecA-F1 and lecA-R1 (Table 2.3) which contained *Nco*I and *Bgl*II sites. The PCR products were analysed by agarose gel electrophoresis and a band corresponding to the expected size was present. The PCR product was then ligated into the TA cloning vector pCR2.1 (Table 2.2, Fig 2.1) as described in Fig 2.9 and Fig 3.2. *E. coli* JM109 cells were transformed with the ligation mixture and spread onto LB amp plates containing IPTG and X-Gal. The plasmid DNA isolated from some of the resulting white colonies (Section 2.6), and screened for an insert by restriction analysis with *Eco*RI. The insert was cut out of the plasmid with *Nco*I-*Bgl*II, gel extracted, and ligated to *Nco*I-*Bgl*II digested, and gel extracted pLecB1 vector. The pLecB1 vector is an expression vector that has been established within the laboratory (Creavin *et al.*, *in prep*) which expresses the *P. aeruginosa* lectin PA-IL as a C-terminally (His)₆ tagged protein. After transformation into JM109 cells (Section 2.12), plasmid DNA

was isolated and the presence of the correct insert was confirmed by restriction analysis. As the *lecA* and *lecB* genes are similar in size, the presence of the correct insert was further confirmed by DNA sequencing. The gene was then expressed in a 100 ml culture (Section 2.14) to ensure the recombinant lectin was soluble in *E. coli* (Fig 3.3).

```

1  ATGGCTTGA AAGGTGAGGT TCTGGCTAAT AACGAAGCAG GGCAGGTAAC GTCGATTATC
61  TACAATCCGG GCGATGTCAT TACCATCGTC GCCGCCGGTT GGGCCAGTTA CGGACCTACC
121 CAGAAATGGG GGCCGCAGGG CGATCGGGAG CATCCGGACC AAGGGCTGAT CTGCCACGAT
181 GCGTTTTGTG GTGCGCTGGT CATGAAGATT GGCAACAGCG GAACCATTCG GTCAATACC
241 GGGTTGTTCC GTTGGGTTGC ACCCAATAAT GTCCAGGGTG CAATCACTCT TATCTACAAC
301 GACGTGCCCC GAACCTATGG CAATAACTCC GGCTCGTTCA GTGTCAATAT TGGAAAGGAT
361 CAGTCCTGA

```

Fig 3.1 *P. aeruginosa* *lecA* coding sequence. Nucleotide sequence AF229814; Annealing sites of primers *lecA*-F1 and *lecA*-R1 (Table 2.3) are indicated with red. There are no *Nco*I, *Bgl*II or *Eco*RI restriction sites within the gene.

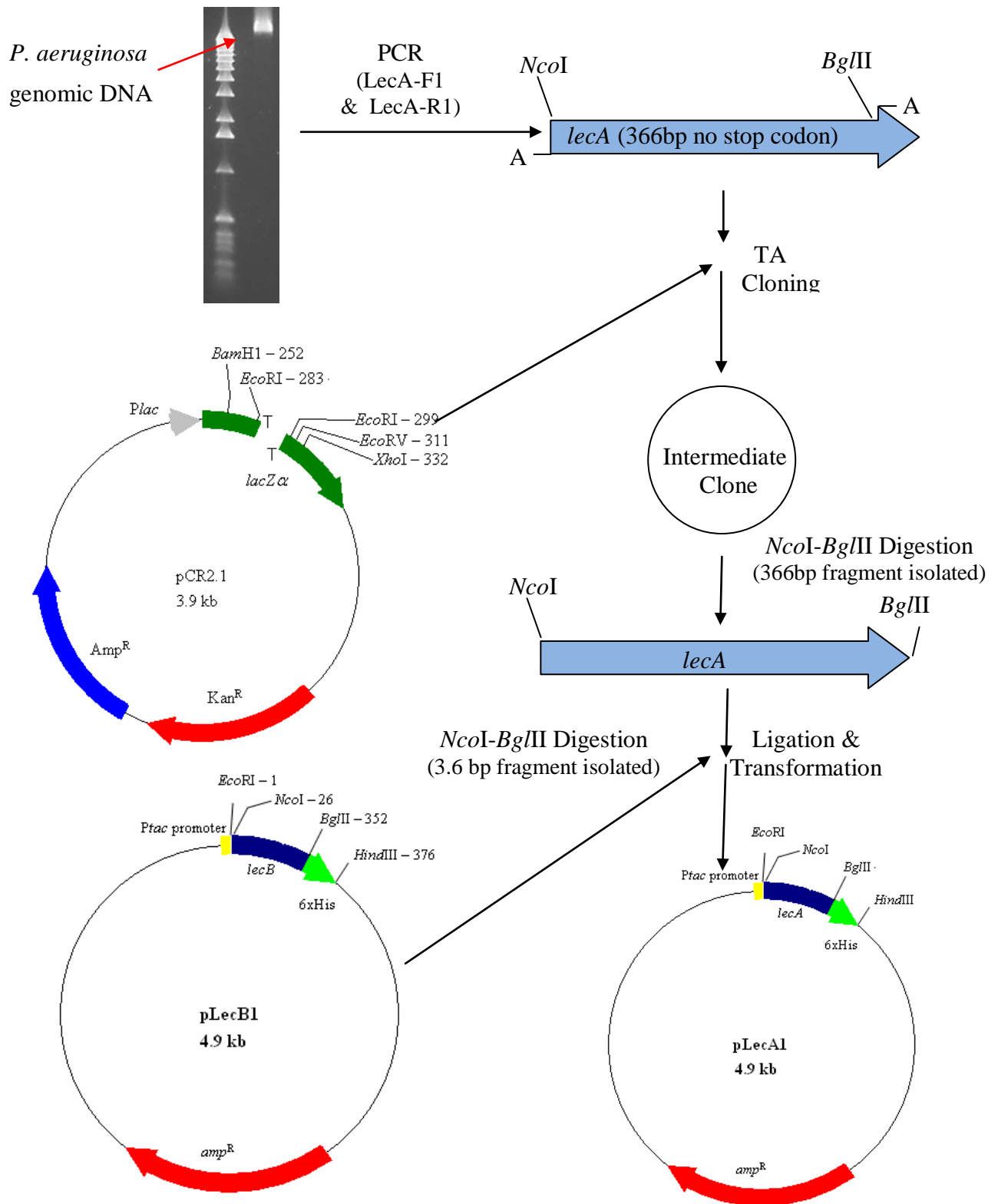


Fig 3.2: Cloning Strategy for pLecA1. Outline of the cloning strategy from genomic DNA to generate pLecA1. The primers LecA-F1 and LecA-R1 amplified the *lecA* gene without the stop codon, and the restriction sites *Nco*I and *Bgl*III at the 5' and 3' ends respectively. The product was ligated into the pCR2.1 vector through TA cloning (Section 2.29), and the insert which was then excised as an *Nco*I-*Bgl*III fragment. The expression vector pLecB1 was digested *Nco*I-*Bgl*III, and the larger band purified by gel extraction (Section 2.10) and ligated with the *lecA* fragment to create pLecA1 vector.

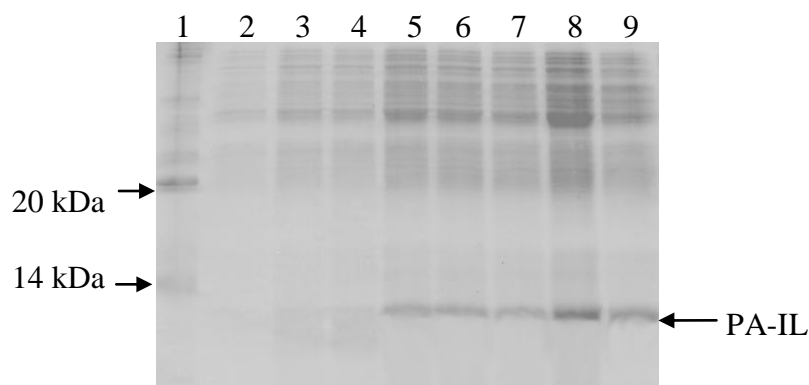


Fig 3.3: Expression of PA-IL in *E. coli* JM109. Analysis by 15% SDS-PAGE (Section 2.18) of the expression of PA-IL from the vector pLecA1 in *E. coli* JM109. Lane 1, Protein molecular weight marker (sizes indicated in Section 2.18); Lane 2, Soluble Cell extract at O.D._{600nm} 0.6; (T₀); Lane 3, T₁ hour; Lane 4, T₂ hour; Lane 5, T₃ hour; Lane 6, T₄ hour; Lane 7, T₅ hour; Lane 8, T₈hour; Lane 9, T₁₈ hours (overnight cell extract).

The gene *lecA* was successfully cloned into the expression vector pLecB1 to create the vector pLecA1 which expressed C-terminally (His)₆ tagged PA-IL. The construct was verified by gene sequencing. The protein was expressed at small scale (Section 2.14) in *E. coli* JM109, and the cell lysate examined by SDS-PAGE (Fig 3.3). An over-expressed protein of the correct size (13.8 kDa) was visible.

3.2.2 Cloning and small-scale expression of *P. luminescens* gene *plu2096* which encodes PL-IL

The *P. luminescens* TTO1 strain was provided by Dr. David Clarke, University of Bath, UK. Genomic DNA was prepared according to the method outlined in Section 2.8, and primers were designed for the amplification of the gene *plu2096* using the TTO1 sequence available through NCBI (Entry No. NC005126). The primers used for the amplification of *plu2096* were plu2096-F and plu2096-R (Table 2.3) which contained *Nco*I and *Bgl*III sites and after amplification with RedTaq polymerase, were ligated into the pCR2.1 cloning vector as was previously done for *lecA* in Section 3.2. This intermediate vector was digested *Nco*I-*Bgl*III, the gene was excised

by gel extraction, and was ligated to an *NcoI*-*BglIII* digested pLecB1 plasmid. As *plu2096* and *lecB* are similar in size, the presence of the correct insert was confirmed by DNA sequencing. As the protein encoded by *plu2096* is PL-IL, the resulting expression vector was named pPL-IL. The gene was then expressed in a 100 ml culture to ensure the recombinant lectin was soluble in *E. coli* (Fig 3.5).

```

1  ATGTCTGATT GGTCAGGAAG TGTTCCAGCT AACGCTGAAA ATGGTAAATC CACAGGACTT
61 ATTCTCAAAC AGGGAGATAC AATCTCTGTT GTTGCTCATG GATGGGTGAA ATATGGCCGT
121 GATAATGTTG AATGGGCAGC ACCTGATGGT CCTGTACCCA ATAATCCACA ACCGTCATCA
181 ATAGCCACTC TCGTCGCTAA GATCGCCAAC AAGAAGTTTG CAATTGGCAA CGGAGTACTT
241 CATAAAACGG TTCCTGTAGA TGGCGAATTA ATACTTTTAT TCAATGACGT ACCGGGTACT
301 TTTGGTGATA ATTCAGGTGA ATTTCAAGTC GAGGTCATAA TAGAGTCTCG ATACTCCCTT
361 TTAAAATAA

```

Fig 3.4 : Sequence analysis of the *P. luminescens plu2096* coding sequence. The 369 bp coding sequence of *plu2096* located within nucleotide sequence NC005126 available on the NCBI database. Annealing sites of primers *plu2096*-F and *plu2096*-R (Table 2.3) are indicated with red. There are no *NcoI* or *BglIII* sites within the sequence.

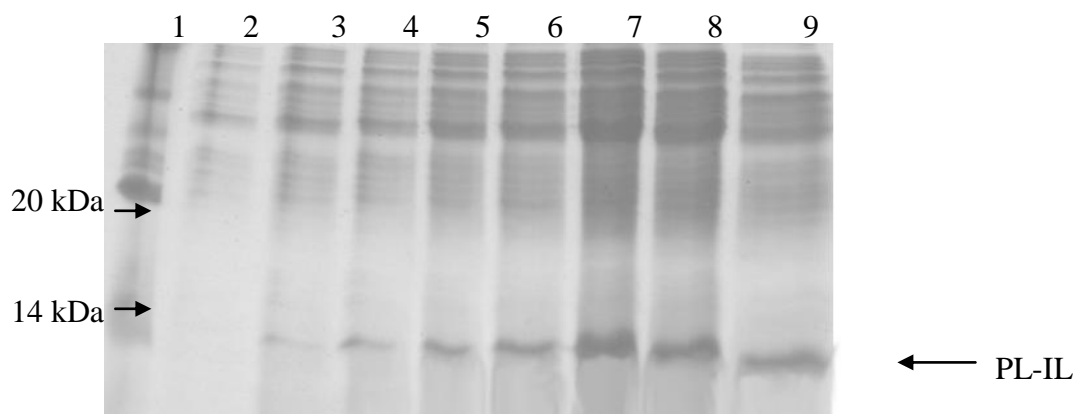


Fig 3.5: Expression of PL-IL in *E. coli* JM109. Analysis by 15% SDS-PAGE of the expression of PL-IL of in *E. coli* JM109. Lane 1, Protein molecular weight marker (sizes indicated in Section 2.18); Lane 2, soluble cell extract at O.D._{600nm} 0.6; (T_0); Lane 3, T_1 hour; Lane 4, T_2 hour; Lane 5, T_3 hour; Lane 6, T_4 hour; Lane 7, T_5 hour; Lane 8, T_8 hour; Lane 9, T_{18} hours (overnight cell extract).

The gene *plu2096* was successfully cloned into the expression vector pLecB1 using the same strategy as outlined in Fig 3.2 to create the vector pPL-IL1 which expressed C-terminally (His)₆ tagged PL-IL. The construct was verified by gene sequencing. The protein was expressed at small scale (Section 2.14) in *E. coli* JM109, and the cell lysate examined by SDS-PAGE (Fig 3.5). An over-expressed protein of the correct size was visible (13.9 kDa).

3.2.3 Mutagenesis of PA-IL, and small-scale expression of the mutated gene encoding PA-IL_{mut1}

As described previously, the lectins PA-IL and PL-IL show high similarity in their amino acid sequences (Section 1.9.2), with the highest identity in the calcium binding site. The non-calcium binding amino acids that contacted the sugar did display some sequence differences however. To determine if these residues played a role in sugar specificity, the *lecA* sequence was altered within the sugar binding loop by site directed mutagenesis, from the amino acids Histidine, Aspartic Acid and Glutamine (PA-IL), to the corresponding amino acids found in PL-IL, namely, Valine, Asparagine and Asparagine. The resulting gene was named *lecA_{mut1}* and the protein named PA-IL_{mut1}. This molecule is essentially the PA-IL lectin, containing the PL-IL sugar binding loop. The procedure used for site-directed mutagenesis is outlined in Fig 3.6. The mutation of *lecA* is outlined in Fig 3.7. The predicted impact to the protein structure was modelled using 3D-structure analysis software, with the theoretical structure of the mutant protein given in Fig 3.8, and the resulting amino acid sequence of PA-IL_{mut1} shown in Fig 3.9.

Having successfully constructed a plasmid expressing the mutated gene, it was then examined whether a soluble protein could be expressed in *E. coli* (Fig 3.10).

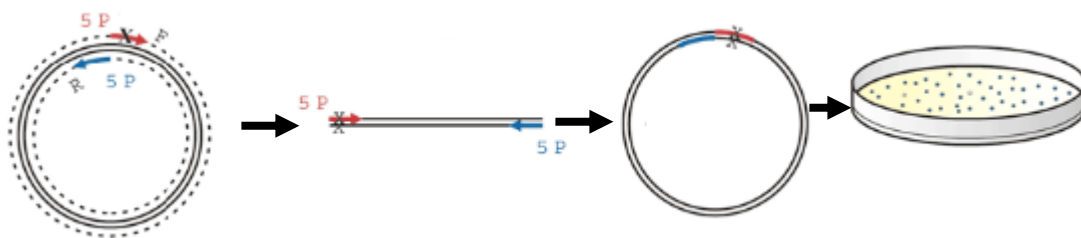


Fig 3.6: Schematic of Phusion™ site directed mutagenesis. Two primers are designed to amplify the entire expression plasmid, one or both of which contains the desired mutation. After PCR (Section 2.30), the phosphorylated primers can be ligated (Section 2.30) and transformed (Section 2.12) directly into an *E. coli* expression strain. (Image adapted from Finnzymes Ltd, Finland). A silent mutation which introduces a restriction site can be incorporated into the junction of the forward and reverse primers to help distinguish between the template plasmid and the resulting mutated plasmid. For the mutation of *lecA*, the template plasmid was pLecA1.

```

atggccttgga aaggtgaggt tctggctaataaacgaagcag ggcaggtaac gtcgattatc
tacaatccgg gcgatgtcat taccatcgtc gccgccggtt gggccagtta cggacctacc
                                gtt  aaca ac                                ac
cagaaatggg ggccgaggg cgatcgggag CTtccgGacC aCgggctgat ctgccacgaT
gagt
gCgttttgtg gtgcgctggt catgaagatt ggcaacagcg gaaccattcc ggtcaataacc
gggttggttcc gttggggttgc acccaataat gtccagggtg caatcactct tatctacaac
gacgtgccccg gaacctatgg caataactcc ggctcgttca gtgtcaatat tggaaaggat
agtc

```

Fig 3.7: Outline of the construction of PA-IL_{mut1}. Representation of the position of the two primers used for the mutation of *lecA*, LecAmut1-F (red) and LecAmut1-R (blue) (Table 2.3) within the *lecA* sequence, which create the mutations H50V, D52N and Q53N within the sugar binding site. The nucleotides to be mutated are denoted in capital letters, with the nucleotides introduced on each primer highlighted in green. At the junction of the forward and reverse primers a silent mutation to introduce the restriction site *MluI* was made. This restriction site was introduced to help distinguish between the template plasmid and the resulting mutated plasmid.

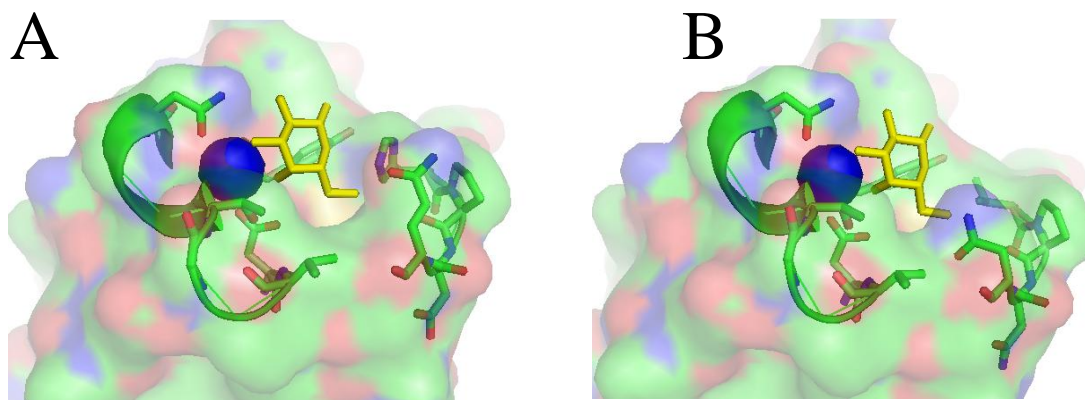


Fig 3.8: Structure of PA-IL binding site and PA-IL_{mut1} binding site. Cartoon representation of the residues involved in calcium and sugar binding within the PA-IL (A) and the PA-IL_{mut1}(B) binding pocket. The calcium ion is symbolized by a blue sphere, with galactose highlighted in yellow. The impact to the shape of the sugar binding loop by the mutation of three amino acids is clearly visible. Image generated using Pymol (Section 2.28), PDB Code 1OKO.

```

      *          20          *          40          *          60          *          80
PA-IL      : MAWKGEVLANNEAGQVTSIIYNPGDVITIVAAGWASYGPTQKWGPQGDREEDQGLICHDAFCGALVMKIGNSGTIPVNT :
PA-IImut1  : MAWKGEVLANNEAGQVTSIIYNPGDVITIVAAGWASYGPTQKWGPQGDREVENNGLICHDAFCGALVMKIGNSGTIPVNT :
            MAWKGEVLANNEAGQVTSIIYNPGDVITIVAAGWASYGPTQKWGPQGDRE P1 GLICHDAFCGALVMKIGNSGTIPVNT

      *          100         *          120
PA-IL      : GLFRWVAPNNVQGAITLIYNDVPGTYGNNSGSFSVNIGKDQS : 122
PA-IImut1  : GLFRWVAPNNVQGAITLIYNDVPGTYGNNSGSFSVNIGKDQS : 122
            GLFRWVAPNNVQGAITLIYNDVPGTYGNNSGSFSVNIGKDQS

```

Fig 3.9: Sequence alignment of PA-IL and PA-IL_{mut1}. Position of the mutations on the amino acid sequence of PA-IL. The conserved proline was not altered as it is conserved through all homologues. The silent mutation that introduced the *Mlu*I restriction site spans C59, H60 and D51.

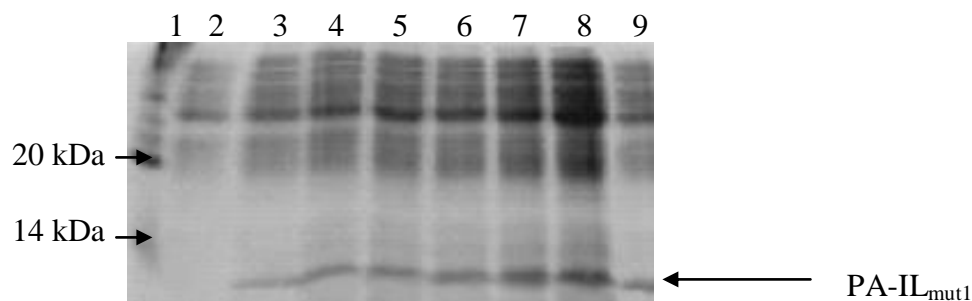


Fig 3.10: Expression of PA-IL_{mut1} in JM109. Analysis by 15% SDS-PAGE (Section 2.18) of the expression of PA-IL_{mut1} into the soluble expression of an *E. coli* JM109 culture. Lane 1, Protein molecular weight marker (sizes indicated in Section 2.18); Lane 2, Soluble Cell extract at O.D._{280nm} 0.6; (T₀); Lane 3, T₁ hour; Lane 4, T₂ hour; Lane 5, T₃ hour; Lane 6, T₄ hour; Lane 7, T₅ hour; Lane 8, T₈hour; Lane 9, Overnight cell extract (T₁₈ hours).

The mutations H50V, D52N and Q53N were successfully introduced into the gene *lecA* to create the gene *lecA_{mut1}* by using the strategy outlined in Fig 3.6 using the plasmid pLecA1 as a template. The resulting expression plasmid was named pLecA4. The construct was verified by gene sequencing. The protein was expressed at small scale (Section 2.14) in *E. coli* JM109, and the cell lysate examined by SDS-PAGE (Fig 3.10). An over-expressed protein of the correct size was visible (13.8 kDa).

3.3 Expression of cloned lectin genes

3.3.1 Sub-cloning of lectin genes into pQE vectors

For the small-scale production of a recombinant lectin that could be purified using IMAC (Section 2.15), the vectors pLecA1, pPL-IL and pLecA4 (Section 3.2) were sufficient. The low level of expression for each protein seen in each case led to the cloning of the genes *lecA*, *lecA_{mut1}* and *plu2096* into expression vectors containing features that would increase yield of recombinant protein. The Qiagen pQE

expression vectors, pQE30 (Fig 2.3) and pQE60 (Fig 2.2), contain various desired elements including an optimised promoter consisting of the phage T5 transcriptional promoter and a *lac* operator sequence, extensive multiple cloning sites, and (His)₆ tag sites at each of the required sites. They also express the β -lactamase gene which confers on the cell resistance to ampicillin. To study the effect of the (His)₆ position however, a panel of differently tagged lectins was required. As the clones already created expressed C-terminally (His)₆ tagged lectins, plasmids that express an N-terminally (His)₆ tagged lectin and an untagged lectin were desired.

To avail of the optimised promoter features on pQE60 (which adds a (His)₆ tag to the C-terminus of a protein), *lecA*, *plu2096* and *lecA_{mut1}* were subcloned into it from pPA-IL, pPL-IL and pLecA4 to determine if a higher yield of protein could be obtained than was previously produced. Conveniently within the pQE-60 MCS site were the recognition sequences for *NcoI* and *BglIII*. As a result, each lectin could be extracted as an *NcoI-BglIII* fragment from pPA-IL, pPL-IL and pLecA4 and inserted into the pQE60 vector to create the vectors pPA-IL60, pPL-IL60 and pPA-IL_{mut1}60.

The pQE30 plasmid, which adds a (His)₆ tag to the N-terminus of a protein, contains within its MCS the recognition sites *BamHI* and *HindIII*. Consequently, the genes *lecA*, *plu2096* and *lecA_{mut1}* and were amplified as *BamHI-HindIII* fragments. The genes *lecA* and *lecA_{mut1}* were amplified using the primers LecA-Fntag, LecA-Rntag, while the gene *plu2096* was amplified using plu2096-Fntag and plu2096-Rntag (Table 2.3). The reverse primer in each case included a stop codon at the 3' end. All lectins were cloned into the pQE-30 (Fig 3.11) vector and the resulting plasmids were named pPA-IL30, pPL-IL30 and pPA-IL_{mut1}30 respectively.

To create untagged PA-IL and PL-IL, the *lecA* and *plu2096* genes were amplified for insertion into the pQE60 vector with a stop codon included in their respective reverse primers. This stopped transcription before the (His)₆ tag located at the 3' end of the gene. In this amplification the same forward primers used in the initial cloning of the respective lectins (LecA-F1, and plu2096-F1) were used in conjunction with a stop codon including reverse primer (LecA-Rstop and plu2096-Rstop). This allowed for the cloning of the genes as *NcoI-BglIII* fragments. The resulting plasmids were

named pPA-ILwt and pPL-ILwt. The sequences of each clone were verified by DNA sequencing. Fig 3.11 shows a schematic of the construction of these plasmids

A summary of the clones created, and the nomenclature that will be used in the remainder of the study is given in Table 3.1

Table 3.1: Summary of lectin nomenclature used in this study.

Gene	Parent Plasmid	Plasmid	Protein	Features
<i>lecA</i>	pLecB1	pLecA1	PA-IL	C-tagged
	pQE30	pPA-IL30	PA-IL30	N-tagged
	pQE60	pPA-IL60	PA-IL60	C-tagged
	pQE60	pPA-ILwt	PA-ILwt	Untagged
<i>plu2096</i>	pLecB1	pPL-IL	PL-IL	C-tagged
	pQE30	pPL-IL30	PL-IL30	N-tagged
	pQE60	pPL-IL60	PL-IL60	C-tagged
	pQE60	pPL-ILwt	PL-ILwt	Untagged
<i>lecA_{mut1}</i>	pLecA1	pLecA4	PA-IL _{mut1}	C-tagged
	pQE30	pPA-IL _{mut1} 30	PA-IL _{mut1} 30	C-tagged

C-(His)₆Tagged (Untagged)

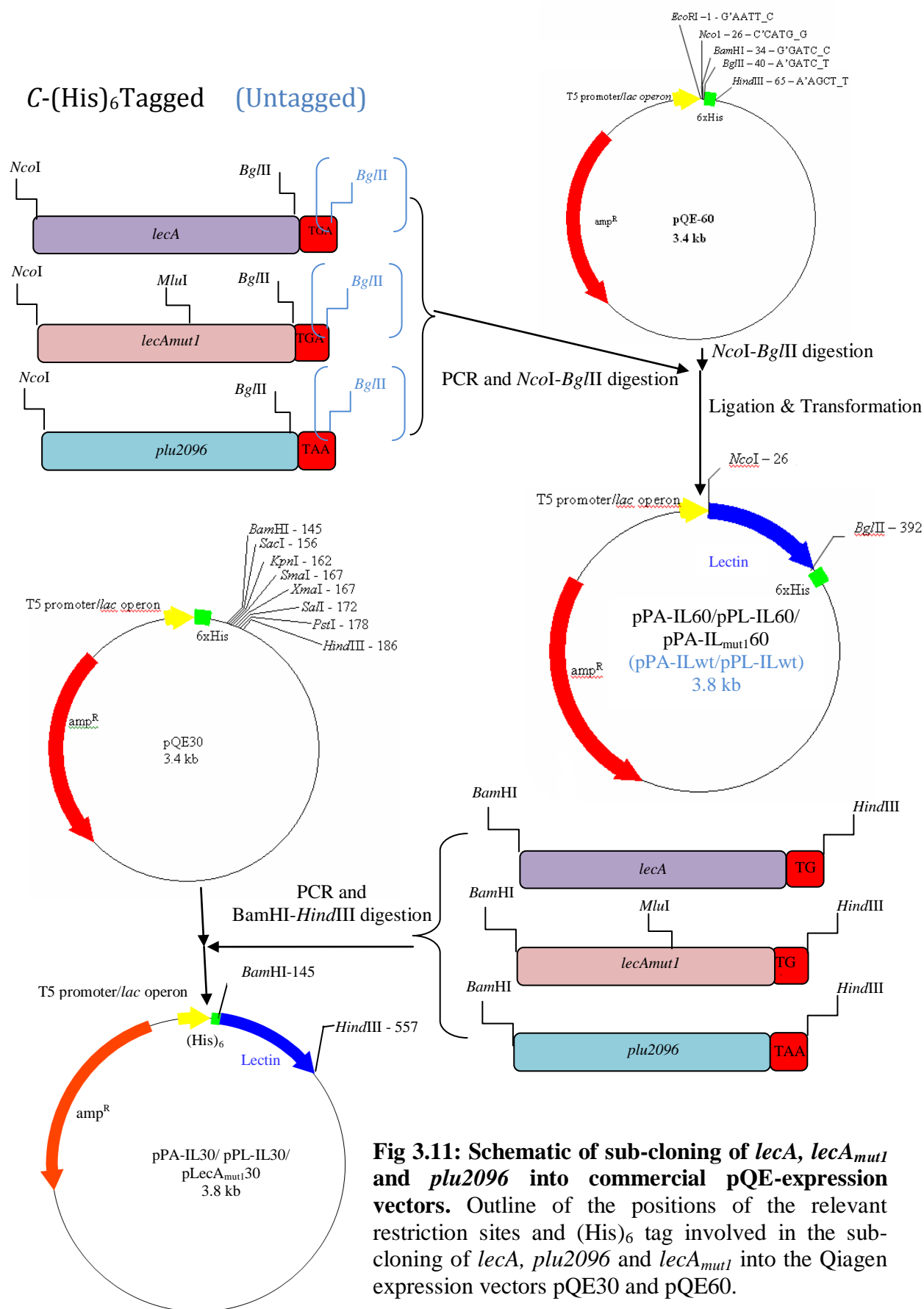


Fig 3.11: Schematic of sub-cloning of *lecA*, *lecA_{mut1}* and *plu2096* into commercial pQE-expression vectors. Outline of the positions of the relevant restriction sites and (His)₆ tag involved in the sub-cloning of *lecA*, *plu2096* and *lecA_{mut1}* into the Qiagen expression vectors pQE30 and pQE60.

3.3.2 Selection of an *E. coli* expression strain for recombinant lectins

E. coli is the most commonly used host for recombinant protein expression. It grows rapidly on inexpensive media and is one of the most highly characterised organisms at a genetic level, with many varying strains commercially available. As the recombinant proteins being investigated (PA-IL and PL-IL) are prokaryotic in source, the question of eukaryotic post-translational modification requirements is not an issue in the choice of an expression host.

The optimal *E. coli* strain for the expression of PA-IL and PL-IL was investigated. PL-IL is a similar sized protein as PA-IL, however its behavior in an *E. coli* expression culture could prove very different to that of PA-IL. As a result, expression conditions needed to be optimized separately. The plasmids pPA-IL30 and pPL-IL30 (Section 3.3.1), which expressed *N*-terminally (His)₆ tagged PA-IL and PL-IL respectively, were transformed into three expression strains to ascertain which was optimal for the expression of recombinant lectin. The three *E. coli* strains used were XL10Gold, BL21, and KRX (Table 2.1).

XL10-Gold is a strain that contains the *lacI*^q allele. This genotype results in a high expression level of the LacI repressor protein, which in turn strongly represses the activity of the *P_{tac}* promoter in the absence of IPTG. BL21 and KRX are commonly used expression strains that are OmpT protease deficient, which should allow for higher recovery of recombinant proteins. Culturing conditions for the three strains were as described in Section 2.14.

The optimal expression strain was determined, by densitometry, by examining the percentage of total soluble protein that was the recombinant protein in the three strains over a six hour culture period (Fig 3.12 and Fig 3.13). The effect of individual clonal selection was then investigated by the colony blot method (Fig 3.15).

After verifying the DNA sequences for each of the clones through gene sequencing, and visual analysis of the purified proteins by SDS-PAGE, the mass of the purified proteins were then examined by electrospray ionization mass spectrometry (Section 2.31). The exact mass (± 2 Da) of the monomeric forms of each lectin could be

evaluated through this method. It was discovered that lectins produced in one strain (XL10Gold), displayed significant product deterioration (Fig 3.16). This was not the case when protein products from other strains were examined, all of which, showed homogenous molecular weight products.

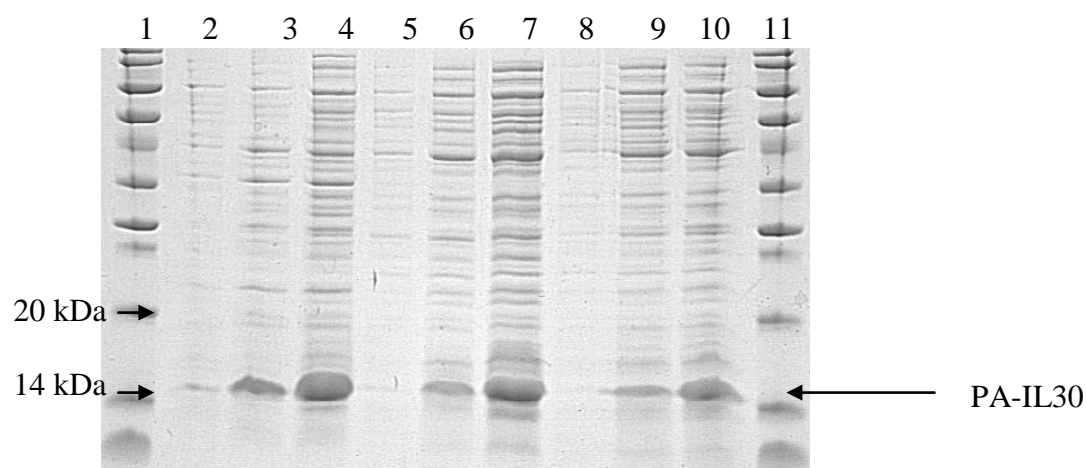


Fig 3.12: Expression of PA-IL30 in BL21, KRX and XL10Gold. Analysis of the expression of PA-IL30 in *E. coli* from pPA-IL30 (Section 3.5). by SDS-PAGE (Section 2.18). LecA30 expression is portrayed as a percentage of total protein, as determined by densitometry. Growth curves for each strain are shown in Fig 21. Lane 1, Ladder (Sizes given in Fig 2.5); Lane 2, XL 10Gold 3.5 hours; Lane 3, XL10Gold 6.5 hours (35%); Lane 4, XL10Gold 9.5 hours (31%); Lane 5, BL21 3 hours; Lane 6, BL21 6 hours (20%), Lane 7, BL21 9 hours (23%); Lane 8, KRX 3.5 hours; Lane 9, KRX 6.5 hours (15%); Lane 10, KRX 9.5 hours (24%) ; Lane 11 Ladder.

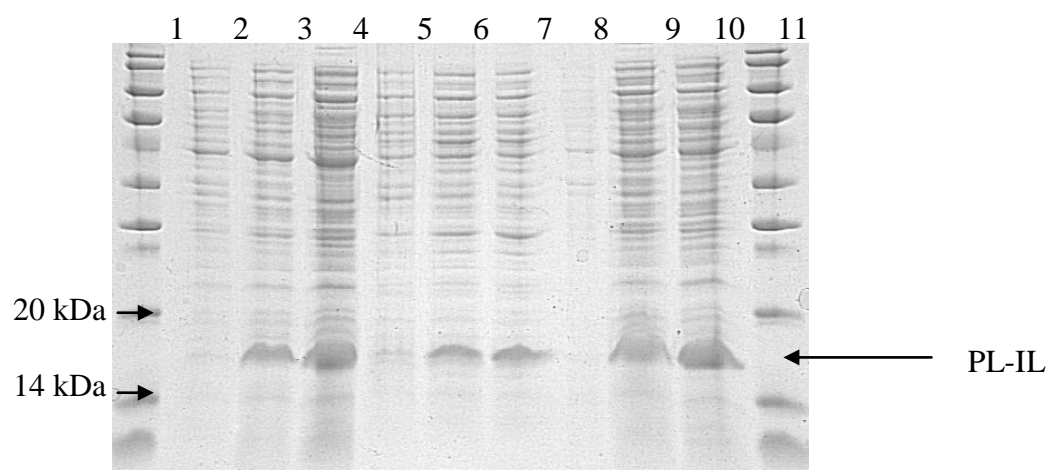


Fig 3.13: Expression of PL-IL in *E. coli* BL21, KRX and XL10Gold. Analysis of the expression of PL-IL30 in *E. coli* from pPL-IL30 (Section 3.3.1) by SDS-PAGE. PL-IL30 expression is portrayed as a percentage of total protein, as determined by densitometry. Growth curves for each strain are shown in Fig 3.21. Lane 1, Ladder (Sizes given in Fig 2.5); Lane 2, XL 10 Gold 3.5 hours; Lane 3, XL10Gold 6.5 hours (17%); Lane 4, XL10Gold 9.5 hours (17%); Lane 5, BL21 3 hours; Lane 6, BL21 6 hours (20%), Lane 7, BL21 9 hours (23%); Lane 8, KRX 3.5 hours; Lane 9, KRX 6.5 hours (16%); Lane 10, KRX 9.5 hours (24%); Lane 11 Ladder.

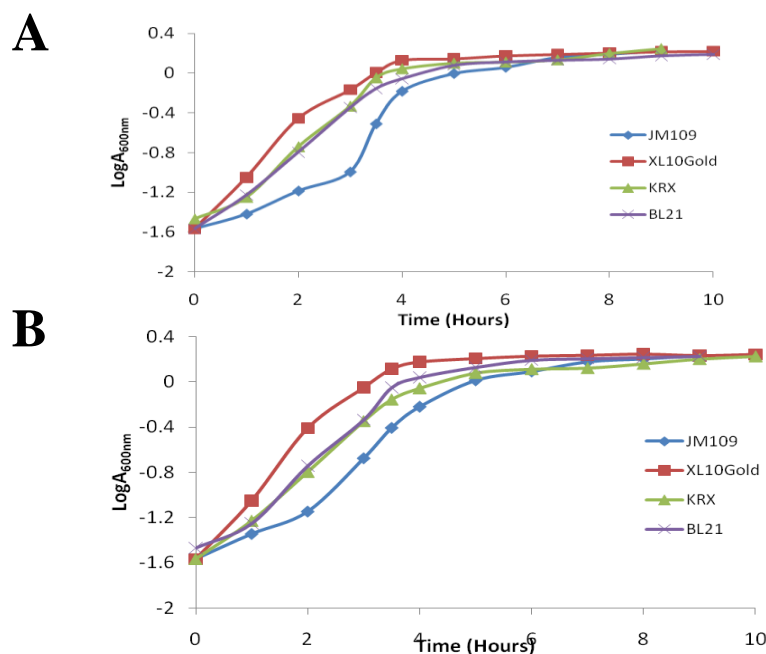


Fig 3.14: Growth curves for the expression of recombinant PA-IL30 and PL-IL30 in *E. coli*. Log absorbance readings recorded at 600nm during the expression of PA-IL30 (Fig A) and PL-IL30 (Fig B) in the *E. coli* strains JM109, XL10Gold, BL21 and KRX. Growth rates for the strains in both experiments represented in terms of doubling time are JM109, 62 min; XL10Gold, 34 min ; KRX, 42 min; BL21, 43 min; and in PL-IL JM109, 57 min, XL10Gold, 31 min, KRX, 40 min, BL21, 40 min.

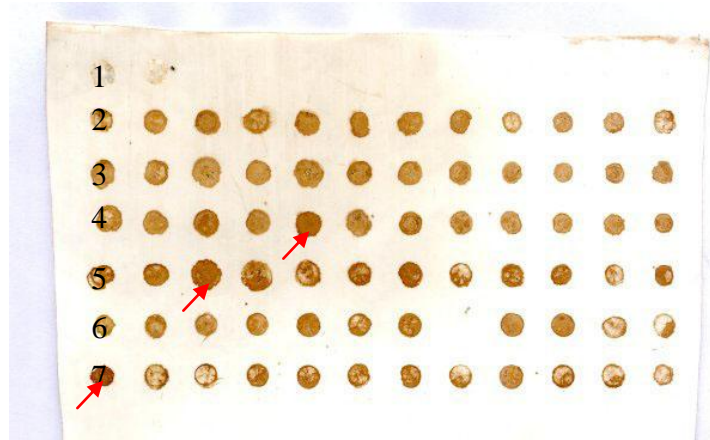


Fig 3.15: Effect of colony selection on lectin expression in *E. coli* KRX. Colony blot of PA-IL constructs expressed in *E. coli* KRX (Section 2.14.2). Row 1, KRX cells containing pQE60; Row 2, PA-IL30 in KRX; Row 3, PA-IL60 in KRX; Row 4, PA-IL_{mut1}30; Lane 5, PA-IL_{mut1}60; Lane 6, PA-IL (Section 3.2.1); Lane 7, PA-ILgfp30 (not included in this thesis). Highlighted with red arrows are individual clones that were shown to express higher levels of protein that their counterparts.

In this optimisation study the PA-IL30 and PL-IL30 molecules were expressed in a number of different *E. coli* strains to determine which strain yielded the highest amount of soluble protein. In a time-course experiment, SDS-PAGE analysis shown that for PA-IL this strain was XL10Gold (Fig 3.12), and for PL-IL the optimal strain was BL21 (Fig 3.13). The growth rates of each strain were also investigated, with XL10Gold cells proving to have faster growth rate than BL21 and KRX, which behaved in a similar manner (Fig 3.14). Upon investigation by ES-MS of purified PA-IL from XL10Gold and PL-IL from BL21, it was discovered that deterioration of PA-IL was significant, which was not seen with PL-IL (Fig 3.16). As the second most efficient expression strain for PA-IL was KRX, this strain was then used to express this molecule. KRX expressed PA-IL did not display any such deterioration. Upon selection of a clone that was deemed to be a successful expressor (Fig 3.15), a glycerol stock was created, which could then be referred to when lectin was desired.

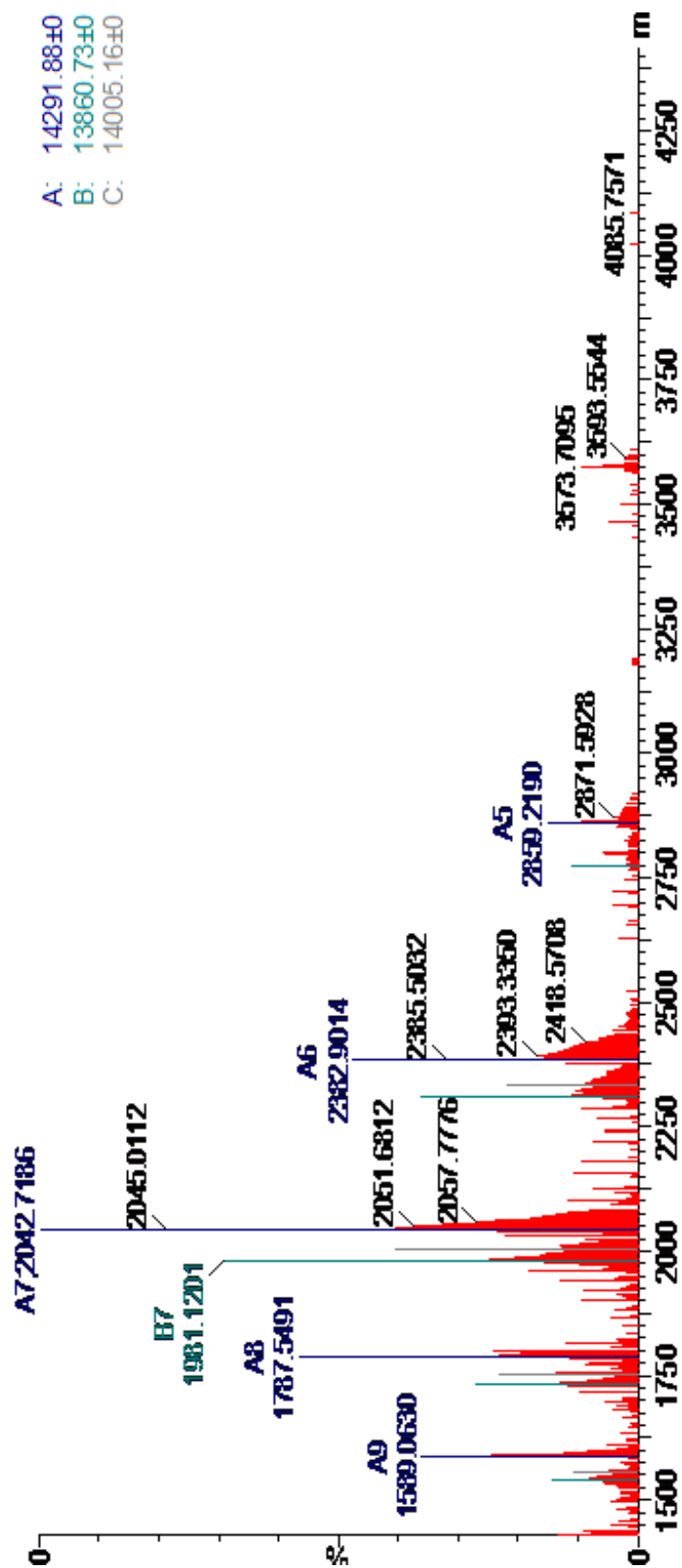


Fig 3.16: Degradation of *E. coli* XL-10Gold expressed PA-ILmut1-30 visualised by ES-MS. Denatured PA-ILmut1-30 at collision energy 60v, with pressure 2.8 mbar in formic acid 1% formic acid (v/v). Three different molecular weight molecules are present with sizes of 14292 kDa, 14005 kDa and 13860 kDa corresponding to the PA-ILmut1-30 molecule minus the *N*-terminal residues methionine (blue), methionine arginine (cyan) and methionine arginine serine (grey) respectively are visible.

3.3.3 Optimization of expression conditions for recombinant lectins

Having elucidated in the previous section that *E. coli* KRX is the optimal strain for the production of PA-IL, and BL21 is the optimal strain for the production of PL-IL the conditions for induction and harvesting of the expressed proteins were investigated.

Both lectins were expressed from plasmids derived from the pQE range of expression plasmids (Section 3.3.2), which are under the control of the P_{TAC} operator system. The tac promoter/operator (P_{TAC}) is one of the most widely used expression systems, and is a strong hybrid promoter composed of the -35 region of the trp promoter and the -10 region of the lacUV5 promoter/operator. Expression of P_{TAC} is repressed by the LacI protein. The lacI^q allele is a promoter mutation that increases the intracellular concentration of LacI repressor, resulting in strong repression of P_{TAC} . Addition of the inducer molecule IPTG inactivates the LacI repressor. Thus, the amount of expression from P_{TAC} is proportional to the concentration of IPTG added: low concentrations of IPTG result in relatively low expression from P_{TAC} and high concentrations of IPTG result in high expression from P_{TAC} . The IPTG concentration was varied in PL-IL expression and the amount of protein product expressed from P_{TAC} in its soluble form examined by SDS-PAGE (Fig 3.17). The effect of increasing IPTG levels on growth rate was also examined (Fig 3.18).

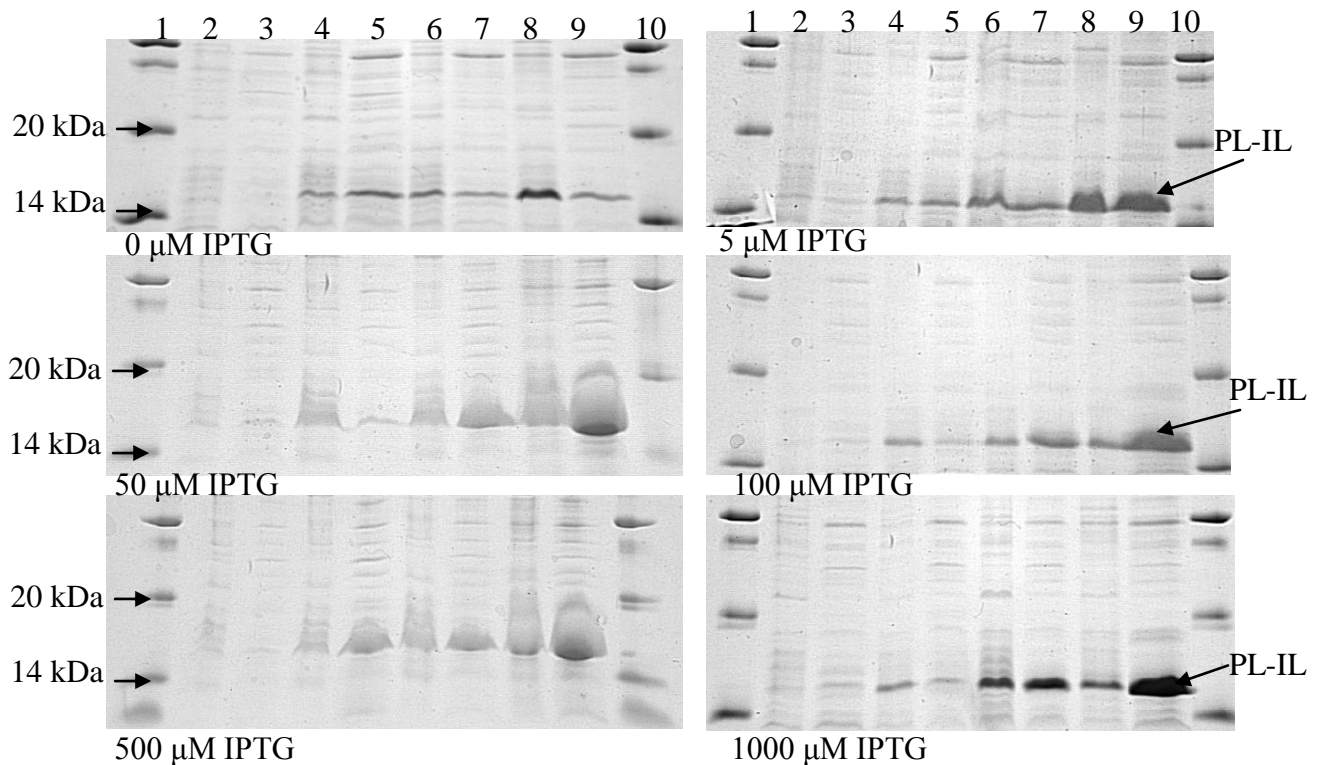


Fig 3.17: Effect of varying IPTG concentrations on expression of PL-IL in BL21. Analysis of increasing IPTG concentration on the expression of recombinant PL-IL in *E. coli* BL21 by 20% SDS-PAGE. Lane 1, Protein marker; Lane 2, insoluble fraction 0 hours; Lane 3, Soluble fraction 0 hours; Lane 4, Insoluble fraction 2 hours; Lane 5, Soluble fraction 2 hours; Lane 6 Insoluble fraction 4 hours; Lane 7, Soluble fraction 4 hours, Lane 8, Insoluble fraction 6 hours; Lane 9, Soluble fraction 6 hours; Lane 10 Protein marker

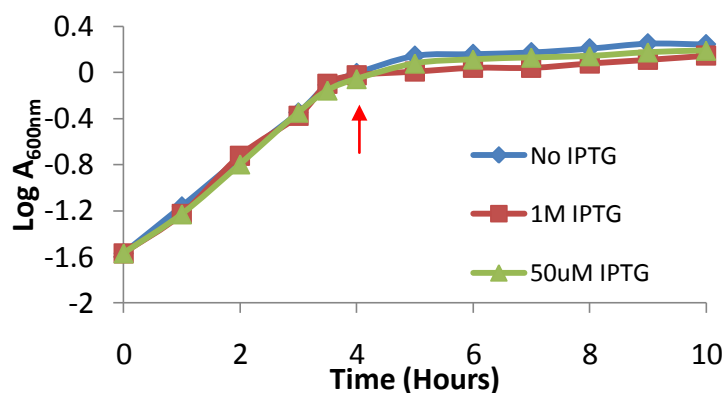


Fig 3.18: Effect of varying IPTG concentrations on growth rate of *E. coli* BL21. Absorbance readings taken during expression of PL-IL30 in *E. coli* BL21 using varying IPTG concentrations (Fig 3.25). The point of induction for the three cultures is indicated with a red arrow. The growth rates for the three cultures (increase in A_{600nm} per hour during exponential phase) was 0.28.

It was deduced from densitometry that there was no significant difference between the amount of soluble protein produced upon induction with 50 μ M IPTG and any higher concentration of the compound (43% of the soluble protein consisted of PL-IL at 50 μ M IPTG, compared to 44% at 500 μ M and 42% at 1 mM). Increasing the IPTG concentration from 50 μ M to 1mM concentration was found not to significantly reduce the growth rate of the cells (Fig 3.18), however, to save on reagents, 50 μ M of IPTG was concluded to be the optimal induction concentration. The harvesting point was then measured by examining the soluble cell extract of an active culture at set points post-induction for both PA-IL in *E. coli* KRX (Fig 3.19) and PL-IL in *E. coli* BL21 (Fig 3.20).

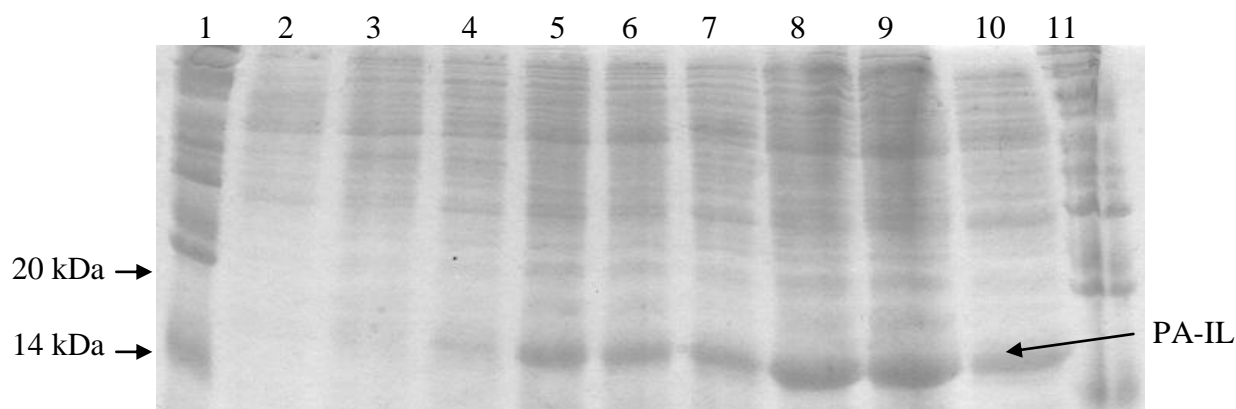


Fig 3.19: Expression of PA-IL in *E. coli* KRX. Analysis of the expression of the protein PA-IL30 from the plasmid pPA-IL30 in *E. coli* KRX by 20% SDS-PAGE. Lane 1, Protein molecular weight. marker; Lane 2, Cell extract at point of induction; Lane 3, 1 hour; Lane 4, 2 hours; Lane 5, 3 hours; Lane 6, 4 hours; Lane 7, 5 hours; Lane 8, 6 hours; Lane 9, 7 hours; Lane 10, overnight culture; Lane 11, Protein molecular weight marker.

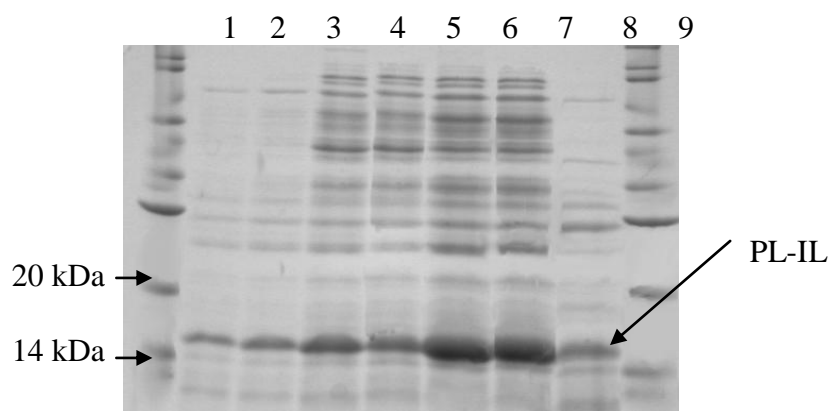


Fig 3.20 Expression of PL-IL30 in *E. coli* BL21. Analysis of the expression of PL-IL from the plasmid pPLIL30 in *E. coli* BL21 by 20% SDS-PAGE. Lane 1, Protein molecular weight marker; Lane 2, Cell extract one hour post-point of induction; Lane 3, 2 hour; Lane 4, 3 hours; Lane 5, 4 hours; Lane 6, 5 hours; Lane 7, 6 hours; Lane 8, 7 hours; Lane 9, 8 hours; Lane 10, overnight culture; Lane 11, Protein molecular weight marker.

Densitometry and SDS-PAGE were used to deduce the optimal harvesting point for the expression of recombinant lectins in *E. coli* KRX and BL21. From SDS-PAGE analysis, it was clear that overnight expression of the lectins significantly reduced the amount of both PA-IL (Fig 3.19) and PL-IL (Fig 3.20) visible in the soluble fraction. At time-point $T_{6 \text{ hours}}$ for both fermentations there was shown not to be the same percentage of product as was visible at time-point $T_{7 \text{ hours}}$ (38% and 38% for KRX; 45% and 44%, for BL21 at $T_{6 \text{ hours}}$ and $T_{7 \text{ hours}}$ respectively) indicating the $T_{6 \text{ hours}}$ time-point was the optimal harvesting point for both strains.

3.4 Purification of recombinant lectins

3.4.1 Purification of untagged lectins

The lectin PA-IL was initially purified using sepharose 4B (Gilboa-Garber *et al.*, 1972). This method exploited the proteins natural affinity for the carbohydrate to separate it from other proteins present in the cytoplasm of *P. aeruginosa*. This was initially attempted after ammonium sulphate precipitation of PA-IL gave only partial purity (Gilboa-Garber, 1972a). A modified version of this technique (Section 2.21) was carried out on the cell extract of a 100ml culture of *E. coli* expressing the lectin from the pPA-ILwt plasmid (Section 3.5). Fig 3.21 shows the SDS-PAGE analysis of the successful purification of untagged PA-IL using sepharose, using galactose to elute bound protein, with an O.D_{280nm} trace of proteins eluted from the sepharose column displayed in Fig 3.22.

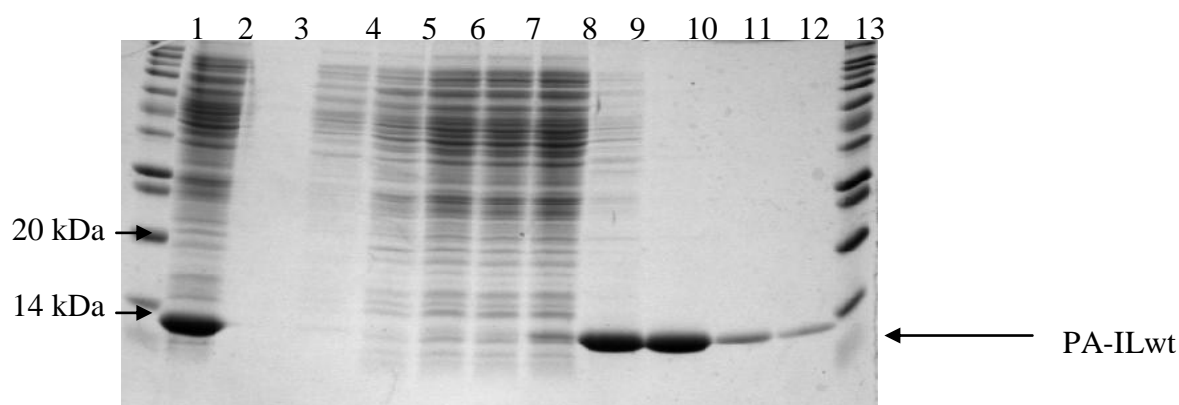


Fig 3.21: Purification of untagged PA-IL by affinity chromatography using sepharose-4B. Analysis by 20% SDS-PAGE of PA-ILwt purification by affinity chromatography using sepharose. Lane 1, Molecular weight marker (sizes given in Section 2.29); Lane 2, crude extract; Lanes 3-8, flow through; Lanes 9-12, 100mM galactose elution; Lane 13, Molecular weight marker.

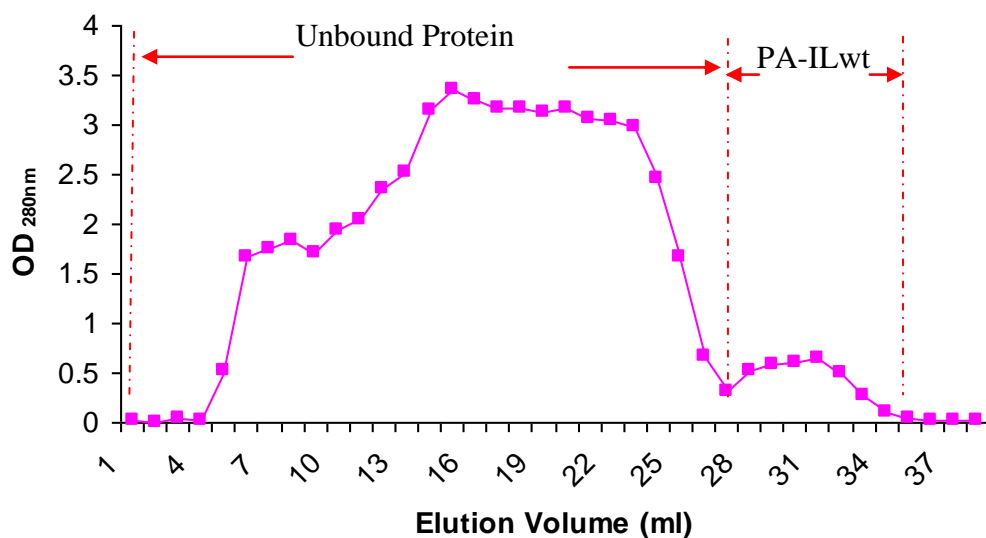


Fig 3.22: Purification of untagged PA-IL using sepharose 4B. Absorption values of the ml fractions from the purification of 10 ml cell lysate from *E. coli* BL21 expressing PA-ILwt from the plasmid vector pPA-ILwt (Section 3.3.1), over a 15 ml Sepharose 4-B column.

It was investigated whether PL-IL and PA-IL_{mut1} had a similar affinity for Sepharose-4B, and whether both lectins be purified from *E. coli* in the same manner. This was done by adding to a sepharose column, the cytoplasmic fraction of an induced culture expressing the lectins PL-ILwt and PA-IL_{mut1}30 (Section 3.3.1) respectively (Fig 3.23A&B), and subsequent elution with galactose or EDTA. The PL-IL and PA-IL_{mut1} molecules could not be purified from *E. coli* cell extract by affinity chromatography using sepharose. It was later shown in Section 4.7 that PA-IL_{mut1} behaved like PA-IL with respect to oligomer assembly, and as a result it was not necessary to purify an untagged version of the molecule.

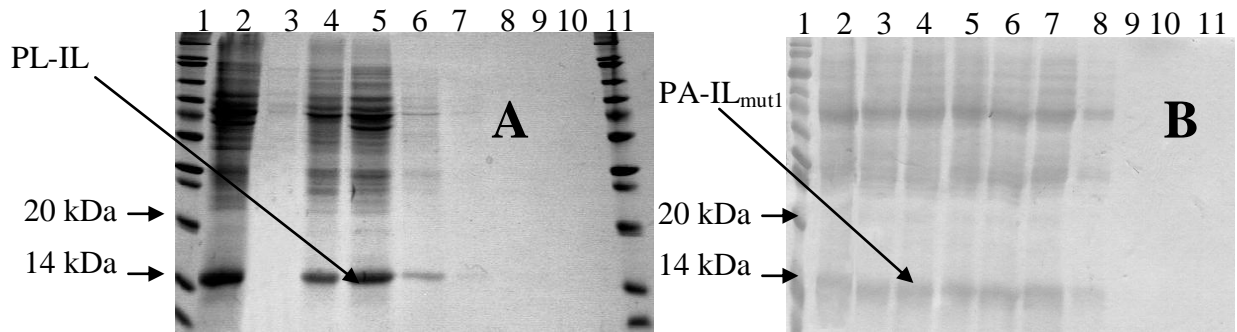


Fig 3.23A&B: Attempted purification of PL-ILwt and PA-IL_{mut1}-30 by affinity chromatography using sepharose 4B. *E. coli* BL21 cell lysate expressing untagged PL-IL from the plasmid pPL-ILwt (A) and C-tagged PA-IL_{mut1} from the plasmid pPA-IL_{mut1}(B), was passed over a Sepharose 4B column. The resulting fractions were analysed by 15% SDS-PAGE. Lane 1, Molecular weight marker; Lane 2, Crude cell extract; Lanes 3- 7, column flow through; Lanes 8-9, 100mM galactose elution; Lanes 10, 20mM EDTA elution; (Lane 11, Molecular weight marker)

PL-IL shares significant homology with PA-IL (Section 1.5), and as a result it was thought that it would display many similar physical characteristics that could be exploited for its purification. Prior to the discovery that PA-IL could be purified by affinity chromatography, the protein was crudely purified from *P. aeruginosa* cell extract using a combination of heating, acidification and salt fractionation (Gilboa-Garber, 1972a). This approach was used to purify PL-IL from cleared cell lysate (Figs 3.24, 3.25 and 3.26).

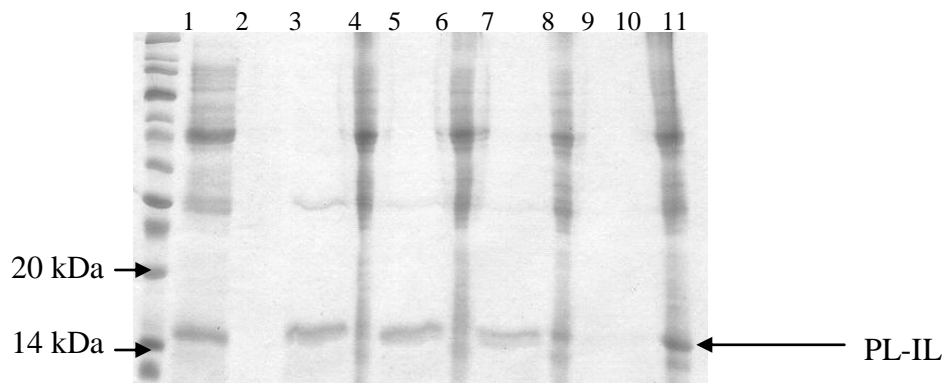


Fig 3.24 Effect of increasing temperature on *E. coli* cell lysate containing over-expressed PL-ILwt. Analysis by 15% SDS-PAGE (Section 2.18) of cell lysate containing over-expressed PL-ILwt heated to varying temperatures and fractionated. Lane 1, Mr; Lane 2, crude cell lysate; Lane 3, 50°C soluble fraction; Lane 4, 50°C insoluble fraction; Lane 5, 60°C soluble fraction; Lane 6, 60°C insoluble fraction; Lane 7, 70°C soluble fraction; Lane 8, 70°C insoluble fraction; Lane 9, 80°C soluble fraction; Lane 10, 80°C, insoluble fraction.

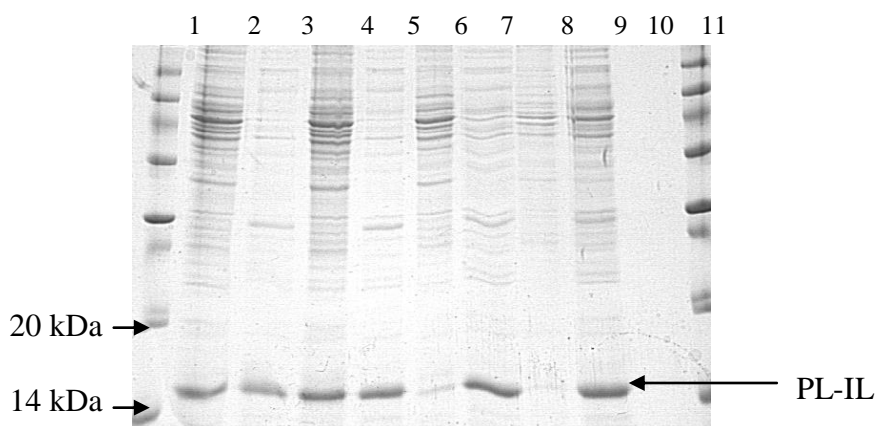


Fig 3.25: Effect of Acetic Acid on Purification by Heating. Analysis by 15% SDS-PAGE (Section 2.18) of cell lysate containing over-expressed PL-ILwt heated to varying temperatures in the presence of 1 mM acetic acid. Lane 1, Mr; Lane 2, Crude cell extract; Lane 3, insoluble cell extract; Lane 4, 50°C soluble fraction; Lane 5, 50°C insoluble fraction; Lane 6, 60°C soluble fraction; Lane 7, 60°C insoluble fraction; Lane 8, 70°C soluble fraction; Lane 9, 70°C insoluble fraction; Lane 10, 80°C soluble fraction; Lane 11, 80°C, insoluble fraction.

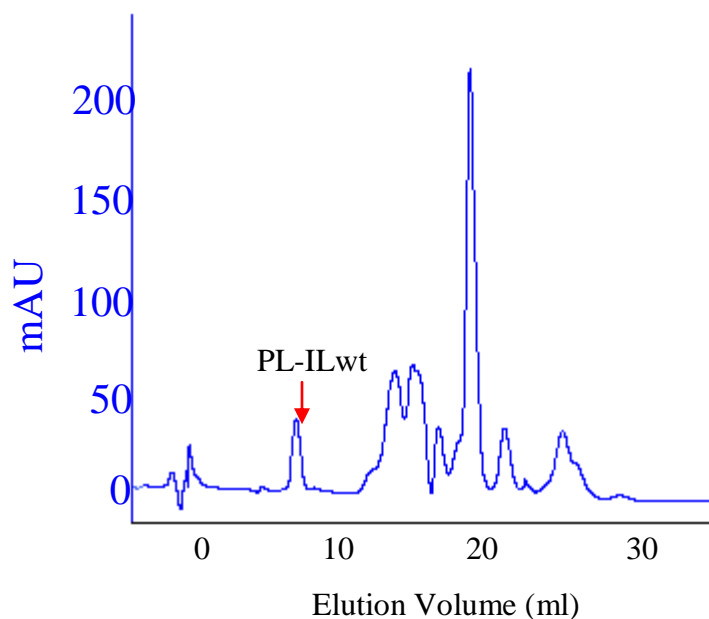


Fig 3.26: Size fractionation of heat treated cell lysate. Elutions measured at OD_{280nm} of elution fractions collected by gel permeation chromatography using the Superdex 75 column matrix (Amersham). of heat treated cell *E. coli* BL21 lysate containing PL-ILwt from the plasmid pPL-ILwt. After SDS-PAGE and Mass-Spectrometry analysis it was confirmed that the highlighted peak represented the purified untagged PL-IL.

PL-IL remained soluble at temperatures up to 70°C, at which point it was the predominant soluble protein. When acid was added to the cell extract, the majority of PL-IL became insoluble above 50°C. As a result, non-acidified, cell extract, which had been heated to 70°C was added to a gel permeation chromatography column, and purified from the remainder of the cellular proteins (Fig 3.26).

3.4.2 Purification of recombinant lectins by IMAC

The purification of histidine tagged lectins through IMAC involves the exploitation of the strong interaction between six histidine residues in the introduced tag and the two ligand binding sites of the Ni²⁺ ion, which has been immobilised on a solid matrix, in this case nitrilotriacetic acid linked to sepharose/sephadex (Ni-NTA). As many cellular proteins will display histidine residues on their surface, several wash

steps are employed using imidazole, the active side-chain of histidine, which displaces loosely bound proteins, and when used in high concentration, displaces the (His)₆ tagged recombinant protein.

Depending on the position of the (His)₆ affinity tag, purification conditions will vary, as the strength of binding between the tag and the nickel matrix is largely dependant on the availability of the histidine residues. The strength of binding will affect the strength of wash tolerated by the interaction, and the point at which within wash gradient the target protein will be eluted. As a result, wash stringencies were investigated for PA-IL30 (Fig 3.29), PA-IL60 (Fig 3.28), PL-IL60 (Fig 3.30) and PL-IL30 (Fig 3.31).

To determine if the matrix composition interferes with the purification of a carbohydrate binding molecule, as it is a sepharose based matrix, a PA-IL30 cell extract was split and passed through two commercial Ni-NTA columns (Qiagen and Amersham) which have slightly different compositions, and an imidazole gradient applied as normal (Fig 3.27 and 3.29).

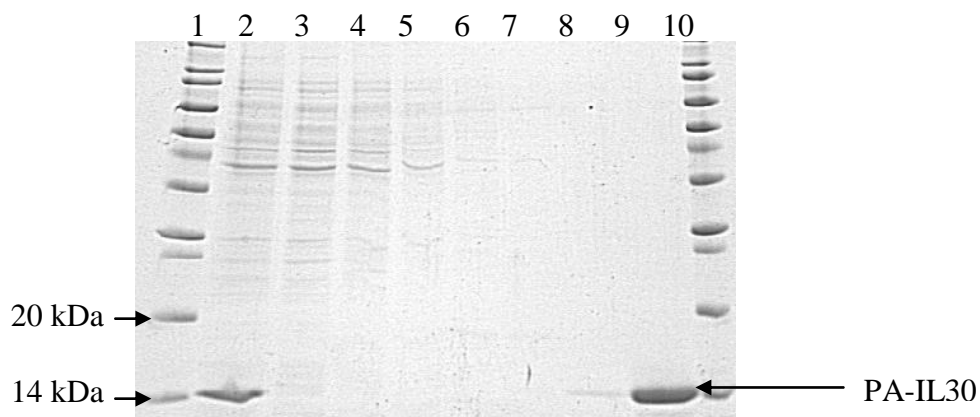


Fig 3.27 Application of cell lysate containing PA-IL30 to Qiagen nickel-NTA resin. Analysis by 20% SDS-PAGE (Section 2.18) of fractions collected from purification on a Nickel-NTA resin (Invitrogen) (Section 2.15.1) of a 100ml *E. coli* cell lysate containing the PA-IL30 protein. Lane 1, Molecular weight marker; Lane 2, crude cell extract; Lane 3, Unbound protein; Lane 4, 20 mM imidazole wash; Lane 5, 40 mM imidazole wash; Lane 6, 60 mM imidazole wash; Lane 7, 80 mM imidazole wash; Lane 8, 100 mM imidazole wash; Lane 9, 300 mM imidazole elution; Lane 10, Molecular weight marker.

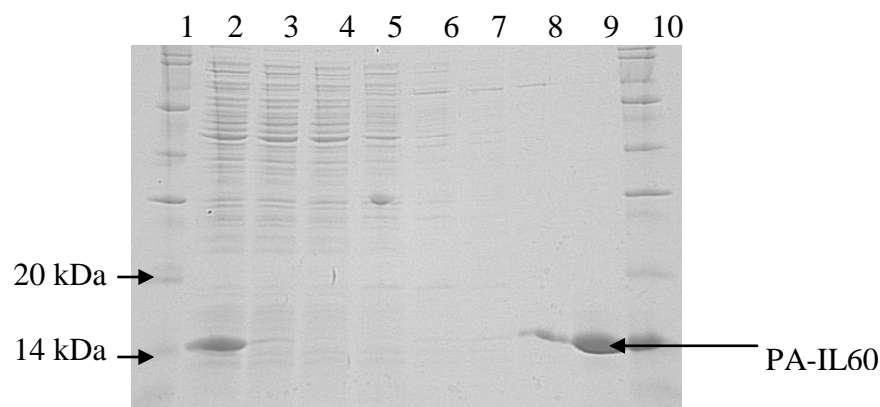


Fig 3.28 Application of cell lysate containing PA-IL60 to Amersham Ni-NTA resin. Analysis by 20% SDS-PAGE (Section 2.18) of fractions collected from purification on an Amersham HisTrap nickel affinity column (Section 2.15.2), of cell lysate from a 100ml culture of *E. coli* cell lysate containing the PA-IL60 protein. Lane 1, Molecular weight marker (Section 2.28); Lane 2, crude cell extract; Lane 3, Unbound protein; Lane 4, 20 mM imidazole wash, Lane 5, 6 mM imidazole wash; Lane 6, 80 mM imidazole wash; Lane 7, 100 mM imidazole wash; Lane 8, 120 mM imidazole wash; Lane 9, 1 M imidazole elution; Lane 10, Molecular weight marker

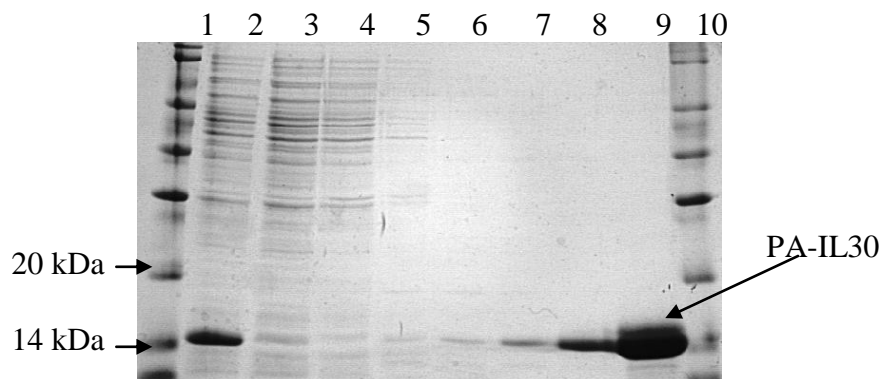


Fig 3.29 Application of PA-IL30 fermentation lysate to Amersham Ni-NTA resin. Analysis by 20% SDS-PAGE (Section 2.18) of fractions collected from purification on an Amersham HisTrap nickel affinity column (Section 2.15.2), of cell lysate from a 100ml culture of *E. coli* expressing the PA-IL30 protein. Lane 1, Molecular weight marker (Section 2.28); Lane 2, crude cell extract; Lane 3, Unbound protein; Lane 4, 40 mM imidazole wash, Lane 5, 60 mM imidazole wash; Lane 6, 80 mM imidazole wash; Lane 7, 100 mM imidazole wash; Lane 8, 120 mM imidazole wash; Lane 9, 300 mM imidazole elution; Lane 10, Molecular weight marker.

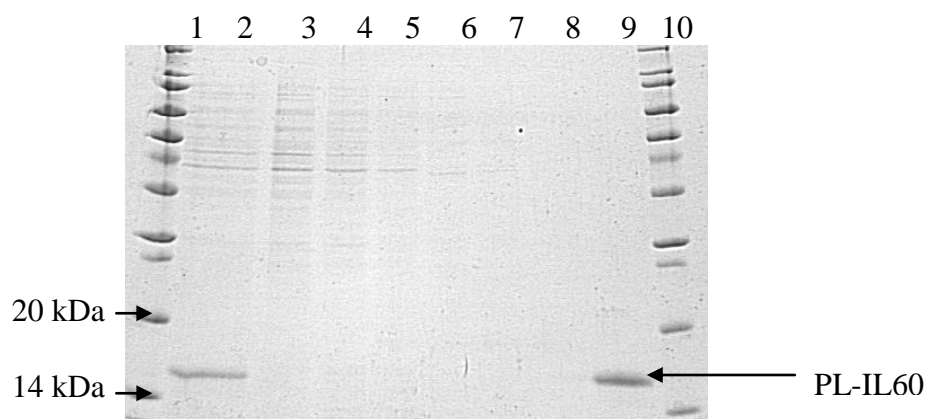


Fig 3.30: Application of cell lysate containing PL-IL60 to Amersham Ni-NTA resin. Analysis by 20% SDS-PAGE (Section 2.18) of fractions collected from purification on an Amersham HisTrap nickel affinity column (Section 2.15.2), of cell lysate from a 100ml culture of *E. coli* expressing the PL-IL60 protein. Lane 1 Mr (Section 2.28); Lane 2, Cleared cell lysate; Lane 3, Unbound column flow through; Lane 4, 40mM imidazole wash; Lane 5, 60mM imidazole wash; Lane 6, 80mM imidazole wash; Lane 7, 100 mM imidazole wash; Lane 8, 120mM imidazole wash; Lane 9, 300 mM imidazole flush; Lane 10, molecular weight marker.

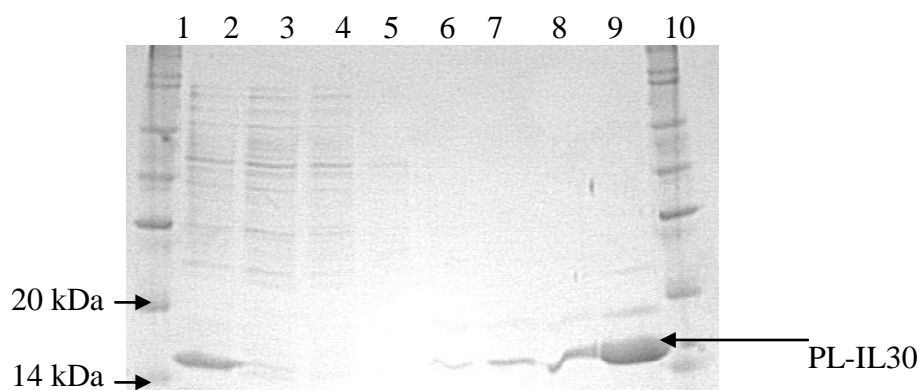


Fig 3.31 Application of cell lysate containing PL-IL30 to Amersham Ni-NTA resin. Analysis by 20% SDS-PAGE (Section 2.18) of fractions collected from purification on an Amersham HisTrap nickel affinity column (Section 2.15.2), of cell lysate from a 100ml culture of *E. coli* expressing the PL-IL30 protein. Lane 1 Mr (Section 2.28); Lane 2, Cleared cell lysate; Lane 3, Unbound column flow through; Lane 4, 40mM imidazole wash; Lane 5, 60mM imidazole wash; Lane 6, 80mM imidazole wash; Lane 7, 100 mM imidazole wash; Lane 8, 120mM imidazole wash; Lane 9, 300mM imidazole elution; Lane 10, molecular weight marker

From SDS-PAGE analysis it was seen that the position of the (His)₆ tag on each lectin affected the purification strategy to be employed. PA-IL60 and PA-IL30 could tolerate imidazole washes of up to 100 mM and 80 mM respectively, before significant amounts of product leached from the column. PL-IL60 and PL-IL30 could tolerate washes of up to 120 mM and 80 mM respectively. It was also found that the composition of the IMAC resin affected the strength of the Ni²⁺ / (His)₆ interaction, as the Qiagen resin was found to bind PA-IL30 more strongly than the Amersham resin (Fig 3.27 and Fig 3.29),

3.4.3 Protein purification using fast protein liquid chromatography (FPLC)

FPLC is a form of chromatography that is commonly employed to separate proteins from complex mixtures found in soluble cell extracts. It was developed in 1982 by Pharmacia and the self-contained system comprises a liquid mobile phase, with a usually solid stationary phase. The solvent velocity is controlled by pumps attached to a reservoir, and the system has the advantage of being compatible with many different column types, which are used depending on what type of separation needs to be exploited. In this case, the column type is an Amersham Ni-NTA column (HisTrap), which is composed of the same elements used in the manual IMAC purification of Section 3.4.2.

Having determined the optimal IMAC purification strategy for lectins using Amersham Ni-NTA resins, this knowledge could be applied to FPLC-compatible Amersham HisTrap columns. For large scale purification of recombinant lectins, this approach would be preferable to manual gravity-flow purification, as there is more control of variables such as flow-rate and pressure as well as having the advantage of being able to monitor properties such as conductivity and absorbance.

Two sample FPLC purification traces are given in Figs 3.32 and 3.33. It is known that PA-IL60 is lost at the wash step of 120 mM imidazole (Fig 3.28), and as a result, a maximum imidazole wash of 100mM was utilised on the FPLC HisTrap column. The protein elution profile can be visualised at OD_{280nm} and can be seen in Fig 3.32, with the corresponding SDS-PAGE analysis of each fraction shown in Fig 3.33. Similarly a PL-IL60 purification strategy using the previously evaluated wash stringency of 120 mM imidazole (Fig 3.30) is shown in Fig 3.34 with the corresponding SDS-PAGE analysis in Fig 3.35.

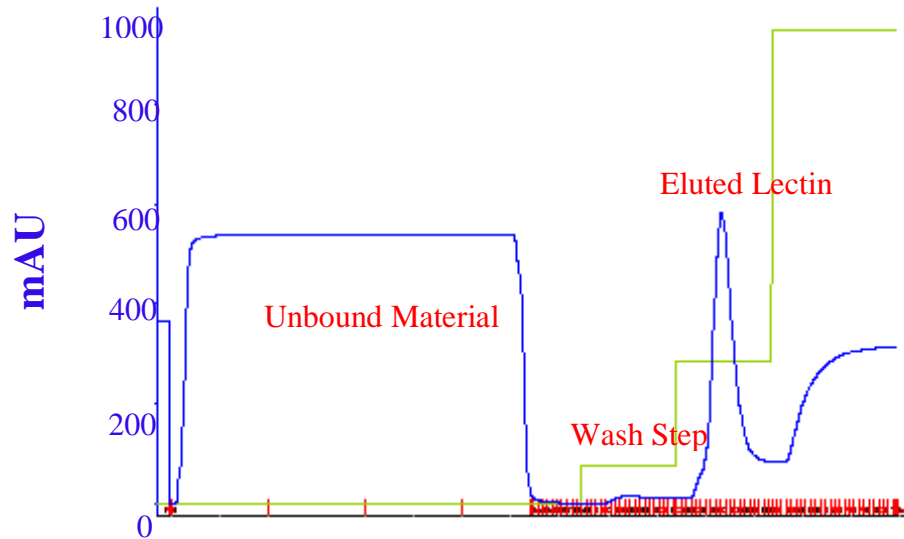


Fig 3.32: Purification of PA-IL60 over HisTrap FF Crude Column by FPLC. Purification profile of cleared cell lysate (Section 2.18), from a 100 ml expression culture of *E. coli* BL21 expressing the protein PA-IL60 from the plasmid pPA-IL60 passed through the Amersham HisTrap FF crude 1 ml column, using FPLC (Section 2.21). The blue line indicates the absorbance at A_{280} , the green line, indicates the percentage of imidazole contained in the running buffer, while the red markers signify the fractionation of the flow-through. The baseline absorbance increasing with respect to imidazole concentration is shown with the brown dashed line.

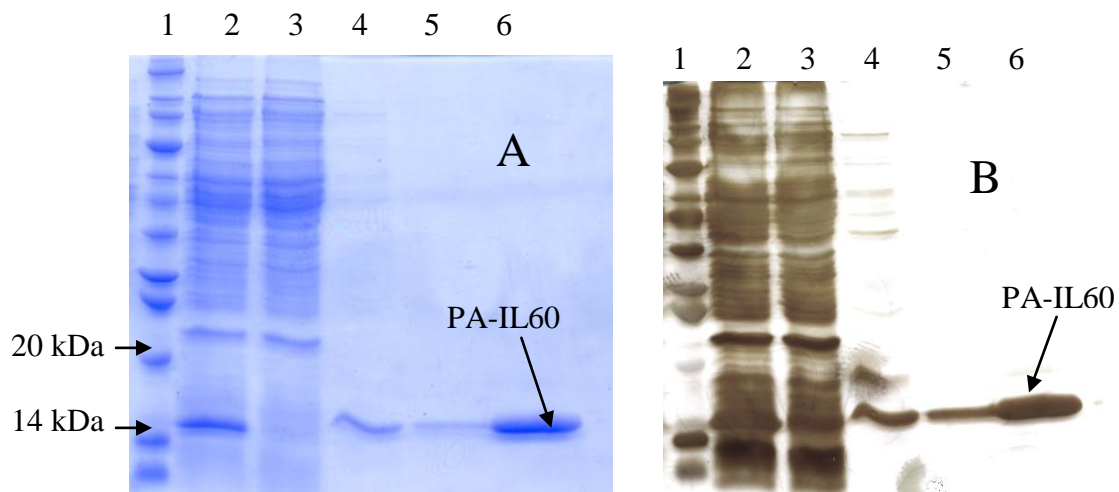


Fig 3.33 Purification of PA-IL60 by FPLC-IMAC. Fractions collected during the IMAC purification of PA-IL60 expressed from pPA-IL60 in *E. coli* KRX analyzed by 15% SDS-PAGE (A) and Silver stained (B) (Section 2.16). Lane 1, Mr; Lane 2, cleared cell lysate; Lane 3, unbound column flow through; Lane 4, 120 mM imidazole wash; Lane 5, 100 mM wash; Lane 6, 300mM imidazole product elution.

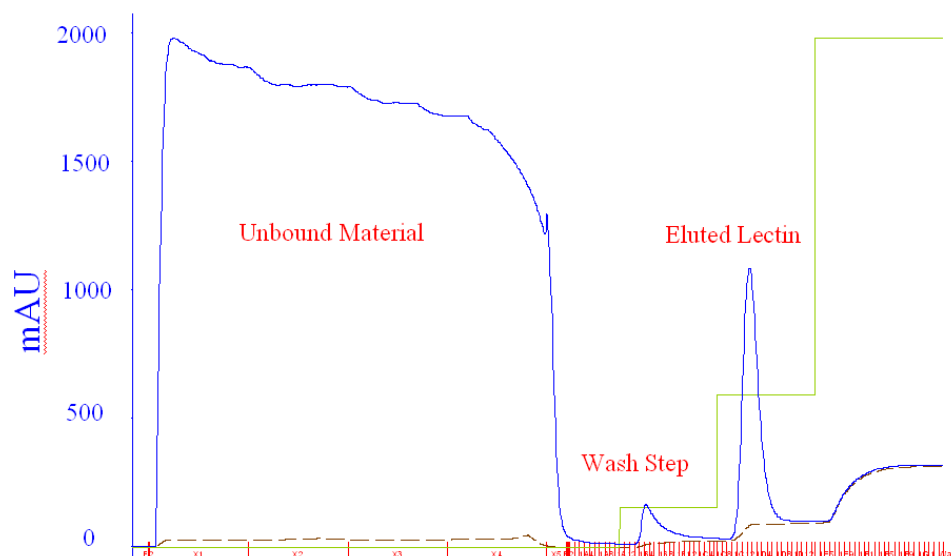


Fig 3.34 Purification of PL-IL60 over HisTrap FF crude column by FPLC. Purification profile of cleared cell lysate from a 500 ml expression culture of *E. coli* BL21 expressing the protein PL-IL60 from the plasmid pPL-IL60 (Section 3.5) passed through the Amersham HisTrap FF crude 1 ml column, using FPLC (Section 2.21). The blue line indicates the absorbance at A_{280} , the green line, indicates the percentage of imidazole contained in the running buffer, while the red markers signify the fractionation of the flow-through. The baseline absorbance increasing with respect to imidazole concentration is shown with the brown dashed line

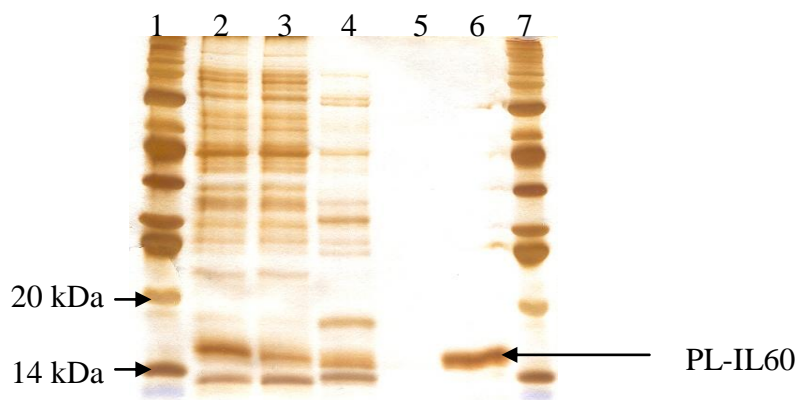


Fig 3.35 Purification of PL-IL60 by FPLC-IMAC. Analysis by silver stained 15% SDS-PAGE (Section 2.18) of fractions collected during the IMAC purification of PL-IL60 expressed from pPL-IL60 (Section 3.3.2) in *E. coli* BL21 Lane 1, Mr; Lane 2, cleared cell lysate; Lane 3, unbound column flow through; Lane 4, 80 mM imidazole wash; Lane 5, 80 mM wash; Lane 6, 300mM imidazole product elution.

It was found that each lectin could be purified successfully on the FPLC using the strategies found optimal in Section 3.4.2. The advantages of purifications using the FPLC include the ability to control parameters such as flow-rate, whereas in manual purification, gravity flow was relied on. Another advantage of the FPLC system is that the quantity of lectin being purified can be accurately determined using the peak height and peak volume recorded for the lectin elution. This saves time by eliminating the requirement for subsequent protein quantification assays. It was also found upon investigation of the purified lectin product by silver-stained SDS-PAGE, that no detectable impurities were present (Fig 3.33 and Fig 3.35).

3.5 Total yields of recombinant lectin

The expression conditions were optimised for recombinant *N*-terminal (His)₆ tagged PA-IL and PL-IL. The optimised PA-IL30 expression conditions were used for all derivatives of PA-IL, including PA-IL_{mut1}. Similarly, the optimal expression conditions found for PL-IL30 were used for the expression of all derivatives of PL-IL. Separate purification strategies were optimised for each recombinant lectin. The average yields for each lectin per g/cell paste after expression and purification optimisation are given in Table 3.2. The average yield is taken after 5 separate fermentations. The g/cell paste value is the weight of cells harvested from a 500 ml culture.

Table 3.2 Protein yields obtained from optimised expression conditions.

Lectin	<i>E. coli</i> Expression Strain	Yield
PA-ILwt	KRX	10-12 mg /5g
PA-IL60	KRX	20 mg /5g
PA-IL30	KRX	25-30 mg /5g
PA-IL _{mut1} 30	KRX	25-30 mg /5g
PL-ILwt	BL21	2-4 mg/5g
PL-IL60	BL21	20 mg/5g
PL-IL30	BL21	20 mg /5g

3.6 Discussion

Extensive optimisation of PL-IL and PA-IL expression was essential due to the high quantity of purified analyte required by analytical methods such as mass spectrometry and gel filtration chromatography. Optimisation of expression conditions is also a pre-requisite for the production of large scale recombinant lectins which would be essential for the application of these novel molecules into formats such as arrays or lectin affinity chromatography columns.

Despite their significant variation on an amino acid level (Fig 1.12), it was hoped that the expression profiles of PA-IL and PL-IL would prove similar in *E. coli* expression systems. PA-IL has already been successfully expressed in *E. coli* (Avichezer *et al.*, 1992), and this result was repeated in Section 3.2, where the addition of the (His)₆ affinity tag to the C-terminus of the lectin by the cloning of the *lecA* gene into the pLecB1 expression vector did not adversely affect the proteins solubility (Fig 3.3).

As the genes encoding PA-IL and PL-IL were similar in size, and lacked the same restriction sites, the same cloning strategy was employed for both lectins (Fig 3.2). The PL-IL protein proved soluble in an *E. coli* expression system (Fig 3.5) with a similar proportion of both JM109 cultures (PA-IL expression and PL-IL expression) being comprised of recombinant lectin.

As the PA-IL molecule has been extensively characterised, with several 3D protein structures available, an attempt was made to mutate PA-IL within the sugar binding site. The PA-IL sugar binding site is comprised of three loops, two involved in the orientation of the essential calcium ion, and one that directly contacts the galactose residue (Section 1.9.3). When the family of PA-IL molecules is examined it is clear that the calcium binding residues remain highly conserved, with only the sugar contacting loop showing considerable variation (Fig 3.36). It was theorized that by mutating this sugar binding loop in PA-IL, the sugar-affinities would be altered in the mutant.

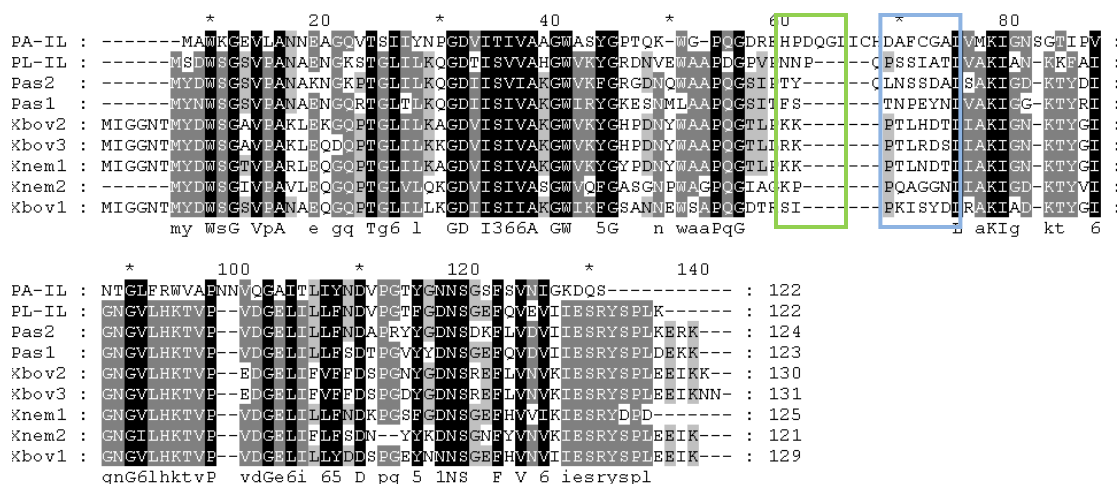


Fig 3.36 Sequence alignment of entire PA-IL-like protein super-family. Amino acid alignment of PA-IL like protein sequences that includes sequences from *Photorhabdus asymbiotica* (Pas), *Xenorhabdus nematophila* (Xnem) and *Xenorhabdus bovienni* (Xbov). Highlighted in green are the amino acid sequences mutated to create the mutant lectin PA-IL_{mut1}. The blue square highlights another variable region positioned close to the sugar binding pocket in PA-IL.

To prove this concept, the PA-IL sugar-contacting loop was mutated to the corresponding loop in PL-IL. A site-directed mutagenesis strategy was employed that consisted of amplifying the entire expression plasmid in a PCR reaction, with the desired mutations encoded on the reverse primer. This method proved successful, with the gene *lecA_{mut1}* created, which expressed a soluble protein named PA-IL_{mut1} (Fig 3.10). Another variable region within the family (highlighted in Fig 3.36) was also mutated using the same strategy, again by changing amino acids in PA-IL, to the corresponding amino acids in PL-IL, however this resulted in an insoluble protein. As a result, the initial mutant remained the focus of this study.

Having achieved the successful expression of soluble recombinant PA-IL, PA-IL_{mut1} and PL-IL in *E. coli*, the genes *lecA*, *lecA_{mut1}* and *plu2096* were cloned into commercial pQE vectors (Qiagen). These are high yield expression plasmids which have a stronger promoter system than pLecB1, a pKK223-3 derived expression plasmid. The optimal *E. coli* expression strain was then determined from three commonly used expression strains, BL21, KRX, XL10Gold. Although most

plasmids were constructed within the *E. coli* strain JM109, this was not included in the expression optimisation study as it is not commonly considered an optimal protein expression strain, and had previously been shown to express soluble protein to a relatively poor level (Figs 3.3, 3.5 and 3.10). Optimisation was not carried out for the expression PA-IL_{mut1}, as it was thought that this would share the same expression characteristics as PA-IL, which only differs to it by three amino acids.

The expression optimisation study looked at the cytoplasmic fraction of three cultures expressing recombinant lectins from the same plasmid (pPA-IL30 or pPL-IL30), and the percentage of total soluble protein that was the desired recombinant protein was calculated. Though the percentage desired protein may be higher in certain strains, the total yield may be higher in another strain, due to a higher cell density and therefore total protein. To combat this, samples to be compared were all taken at the same O.D. rather than the same time-point, as the strains grew at different rates (Fig 3.14). However, even if the same cell mass is present, inefficient cell lysis methods may result in more soluble protein being extracted from one sample than another. Sonication is not the most efficient lysis method that exists, however the other alternative available, freeze-thaw, requires lysozyme which is the same size as the recombinant lectins (14.6 kDa), and so would complicate any SDS-PAGE analysis. As a result, the BCA assay (Section 2.17) was used to estimate the total amount of protein present in each sample, which was then diluted to the same value prior to loading on SDS-PAGE.

The optimal *E. coli* expression strain for PA-IL was determined to be XL10Gold, and for PL-IL, BL21. These recombinant proteins were then purified by IMAC (Section 3.10) and the samples analysed by mass spectrometry, to ensure no protein integrity. However, after mass spectrometry analysis it was discovered that because XL10Gold is not a protease deficient strain, a significant amount of proteolytic degradation of protein product was occurring. This degradation was not visible by SDS-PAGE analysis. A sample trace of PA-IL_{mut1}30 degradation after purification is provided in Fig 3.16. Usually, this would be countered by purifying at 4°C, except when large-scale purification strategies employing the FPLC were required, and this was not an option. Protease inhibitors were also considered, however due to the

hazardous nature and expense of these molecules, especially in a large-scale protein production environment, the protease deficient *E. coli* KRX strain was selected for the expression of PA-IL. Protein products purified from both BL21 and KRX did not display the same proteolytic degradation of product. The total percentage protein yield for this strain was less than XL-10Gold, but total quantity was sacrificed in favour of protein quality.

Having determined the optimal strain for the expression of recombinant PA-IL was KRX, and for PL-IL was BL21, the colony blot method was utilised to distinguish between clones that have high expression rates, as freshly transformed colonies can possess highly varying expression rates. An example dot-blot is given in Fig 3.15, where the majority of colonies expressed the same level of protein, with only a few possessing significantly higher or lower expression rates. This experiment has its limitations however, as it does not distinguish between soluble and insoluble protein. As the chosen strains were BL21 and KRX, with KRX being a derivative of BL21 (Table 2.1), expression conditions were optimised for BL21, assuming that they would transfer to KRX. Induction by IPTG is routinely done at 50 μ M concentration once the O.D_{600nm} reaches 0.5-0.6. However, it is known that PA-IL has a very high affinity for hydrophobically derivatised galactose molecules (Section 1.9.3), an example of which is IPTG. As a result, the effect of induction with varying levels of IPTG was investigated by looking at the percentage of soluble and insoluble product produced at certain stages of the fermentation. As expected, 0 μ M IPTG still allowed the expression of low levels of product. The BL21 strain is known to be a leaky expressor, as unlike KRX and XL10Gold, it does not contain the *lacI*^q mutation, which results in an overproduction of the Lac repressor protein. As a consequence, expression can occur without the addition of the repressor binding molecule IPTG. It was concluded that increasing the IPTG concentration did not result in a greater amount of recombinant protein being expressed to the soluble fraction (Fig 3.17), and that dramatically increasing the IPTG concentration adversely affected the cells growth rate (Fig 3.18), and would have a negative impact on protein yield as a result. It was also determined that the optimal harvest time for fermentation in both *E. coli* KRX and BL21 was 6 hours post-induction (Fig 3.19 and Fig 3.20) as the 7 hours

and overnight samples showed no greater percentage increase in soluble recombinant lectin. It is possible that at this point aggregation of the product within the cytoplasm results in a limit to how much lectin can be produced.

Having evaluated the conditions for the optimal expression of the recombinant proteins within the cell, a purification strategy was then investigated to recover the maximum amount of recombinant protein from the soluble cell extract whilst maintaining the highest level of purity.

PA-IL, as expected, bound to sepharose, and eluted with galactose, and as a result, could be purified from *E. coli* cell extract by affinity chromatography (Fig 3.21). Without prior knowledge of the sugar binding affinities for PL-IL and PA-IL_{mut1}, it was unknown whether the same strategy could be employed for these lectins. They did not bind sepharose (Fig 3.23), indicating either an alternative binding specificity, or a greatly reduced affinity to sepharose compared to that of PA-IL. Rather than cloning a cleavable (His)₆ affinity tag, which would require optimization of the proteolytic cleavage step, PL-IL was purified by a crude mixture of gel filtration and heating. Though acidification was reported to aid in PA-IL purification, it resulted in the loss of significant amounts of PL-IL. However, the considerable stability of the PL-IL structure is seen when the majority of other cellular proteins are denatured, leaving PL-IL soluble. Gel filtration was then used as a polishing step in the purification.

IMAC purification strategies were employed for the purification of (His)₆ tagged lectins. PA-IL30 and PA-IL60 could tolerate imidazole washes up to 80 mM and 100 mM respectively, with PL-IL30 and PL-IL60 tolerating 80 mM and 120 mM strength washes respectively. These optimised conditions were transferred to the automatic FPLC system, which would be employed in the large-scale manufacture of any recombinant lectins. The high level of purity obtainable by solely using IMAC can be seen by the highly sensitive silver stain method, whereupon, no impurities were detectable (Fig 3.33 and Fig 3.35).

4.0 Structural Analysis of Recombinant PA-IL, PA-IL_{mut1} and PL-IL

4.1 Introduction

The expression of recombinant lectins has been reported previously in *E. coli* where it has been shown biologically active plant lectin may be produced (Section 1.8.5). This is a useful alternative to traditional lectin production that required large volumes of source material. For example, the SALT lectin from *Oryza sativa* was expressed and purified from *E. coli* (Branco *et al*, 2004). Previous studies required large volumes of rice plant roots to be used as the protein source.

The effects of affinity tags, incorporated into the recombinant lectins produced in these *E. coli* expression systems are less well characterised. However initial work has been undertaken. For example, PTA agglutinin from *P. ternata* which was expressed in *E. coli* with a (His)₆ affinity tag which was found to be biologically active. Good yields comparable to those seen in Chapter 3 of this thesis were also achieved (Lin *et al*, 2003). Another lectin purified in this way was the galactophilic lectin from *Sarcocystis muris*, which was found to be fully biologically active after the addition of the affinity tag (Klein *et al*, 1998). In this chapter, the effects of the addition of affinity tags to the physical properties of the recombinant lectins PA-IL and PL-IL were examined.

4.2 Biochemical properties of native PA-IL and PL-IL

Previous studies have shown that native PA-IL consists of 121 amino acids, with amino acid sequencing determining that one sub-unit has a M_w of 12,754 Da and a pI of 4.94 (Avichezer *et al*, 1992). Crystallographic data indicates the formation of a tetramer in the native state containing a central hydrophobic pocket and four exposed sugar binding sites, one on each sub-unit (Karaveg *et al*, 2002; Cioci *et al*, 2003). This will be discussed further in Section 1.9.3.

PA-IL shows significant sequence homology to PL-IL (Section 1.9.2), and using PA-IL as a template, the amino acid sequence for PL-IL was submitted to the 3D-

Jigsaw Comparative Protein Modeling web server (Section 2.30), which provides an automatic protein homology modeling service (Bates *et al.*, 2001). The predicted structure for PL-IL is shown in Fig 4.1 where it is compared with the structure of its homologue, PA-IL. Significant similarities are visible throughout the core of the molecule with many sheets and loops displaying an identical structure. Some areas of variation include the extremities of the molecule which include the sugar binding loops and the regions that play a role in oligomerisation in PA-IL.

With respect to PL-IL, there has been no reported investigation of the physical structure other than the *pI* being reported as 5.06 and the relative molecular mass at 12.96 kDa (Turlin *et al.*, 2006).

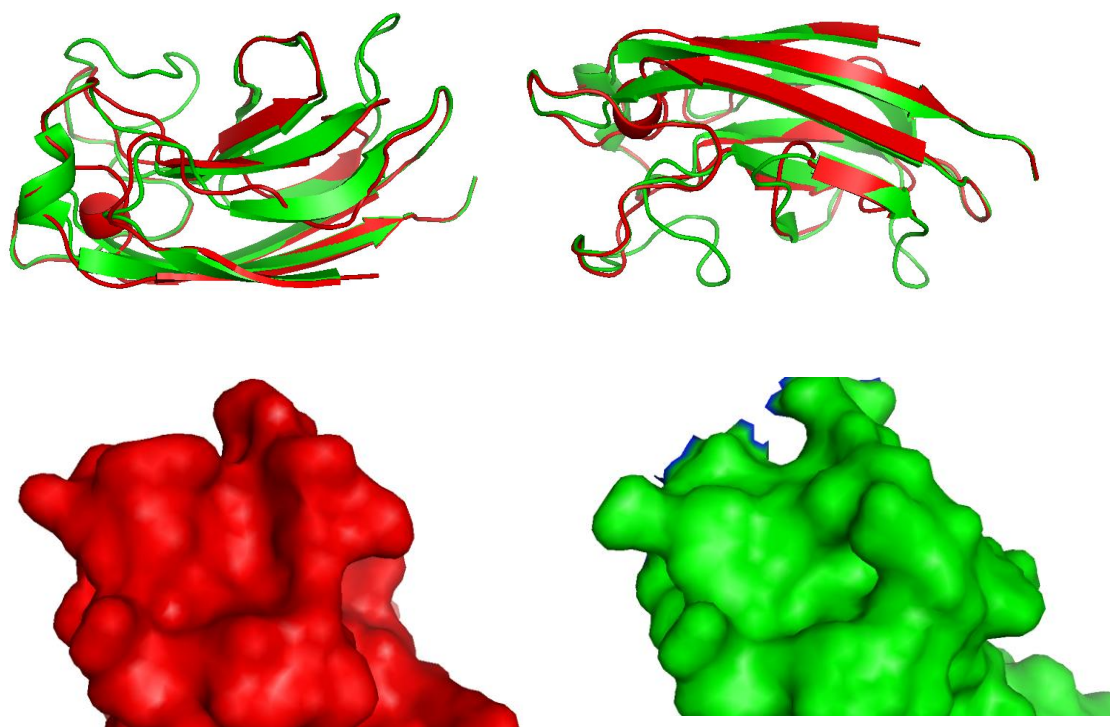


Fig 4.1 Monomeric structure of PL-IL compared to PA-IL as predicted by the 3D-Jigsaw Protein Comparative Modelling Server. Cartoon images of the PA-ILwt monomer (green) and the PA-IL_{mut1} (red) monomer at both the backbone and protein surface levels. Image is displayed using Pymol (Section 2.30) PDB Code 1L7L.

4.3 Construction of a standard curve for lectin molecular weight determination by gel permeation chromatography

To investigate the quaternary structure of the recombinant lectins gel permeation chromatography (GPC) was employed using the superdex 75 gel filtration matrix. This technique separates proteins based on the retardation of lower relative molecular mass proteins within porous beads and the passage of higher relative molecular mass proteins, as they are forced through a packed column. The larger proteins will pass through the column more rapidly than smaller proteins , and their exact size can be elucidated by comparison with a set of standards used to create a standard curve (Fig 4.2 and 4.3). The standards included bovine serum albumin, bovine carbonic anhydrase, ovalbumin, myoglobin and cytochrome C.

Equation 4.1 was used to create the standard curve

$$\text{Equation 4.1} \quad K_a = \frac{V_t - V_e}{V_t - V_o}$$

In this equation V_t represents the fill volume, V_o , the void volume and V_e the elution volume of the protein standard. These were then plotted against the log of the MW to create the graph in Fig 4.2. The elution of each protein on the column can be seen in Fig 4.3.

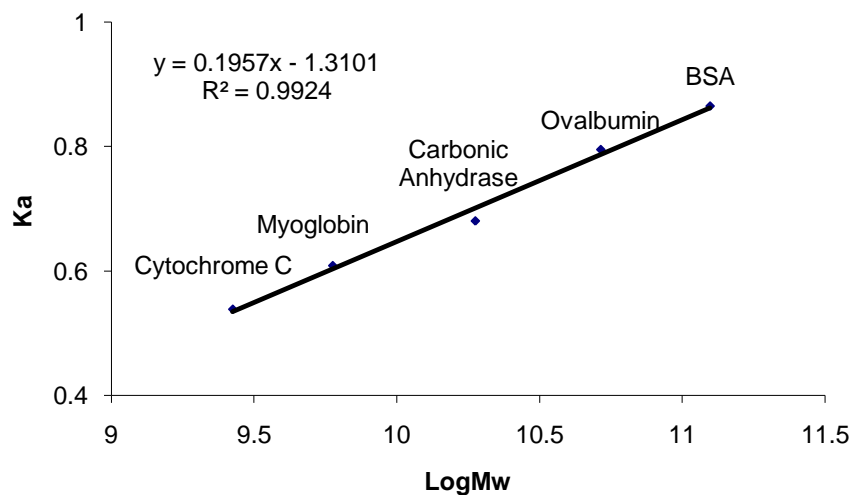


Fig 4.2 Development of size exclusion chromatography standard curve. The protein molecular weight standard curve was created using size exclusion chromatography, with the Superdex 75 10/300 High Performance Column (GE Healthcare) (Section 2.22). Void volume (V_o) of 7.325 ml was determined from the elution volume of blue dextran, and fill volume (V_i) was 18.755 ml, determined with acetone. The elution volumes of protein standards were as follows; BSA, 8.871 ml; Carbonic Anhydrase, 10.98 ml; Ovalbumin, 9.67 ml; Cytochrome C, 12.601 ml; Myoglobin, 11.80 ml. These values were used to construct a plot of K_a versus log MW. This could be in turn used to calculate the relative MW of all lectin samples that were eluted in this range.

Table 4.1 Construction of a protein molecular weight standard curve for the Superdex 75 high performance column (GE Healthcare) at pH 7.2 in PBS.

Standard	Elution Volume	Expected MW	logMW	K _a	Actual MW	%Error
Cytochrome C	12.601 ml	12400	9.43	0.54	12652	2.04
Myoglobin	11.8 ml	17600	9.78	0.61	18100	2.85
Carbonic anhydrase	10.98 ml	29000	10.28	0.68	26115	-9.95
Chicken albumin	9.67 ml	45000	10.71	0.79	46908	4.24
BSA	8.871 ml	66000	11.10	0.86	67046	1.59

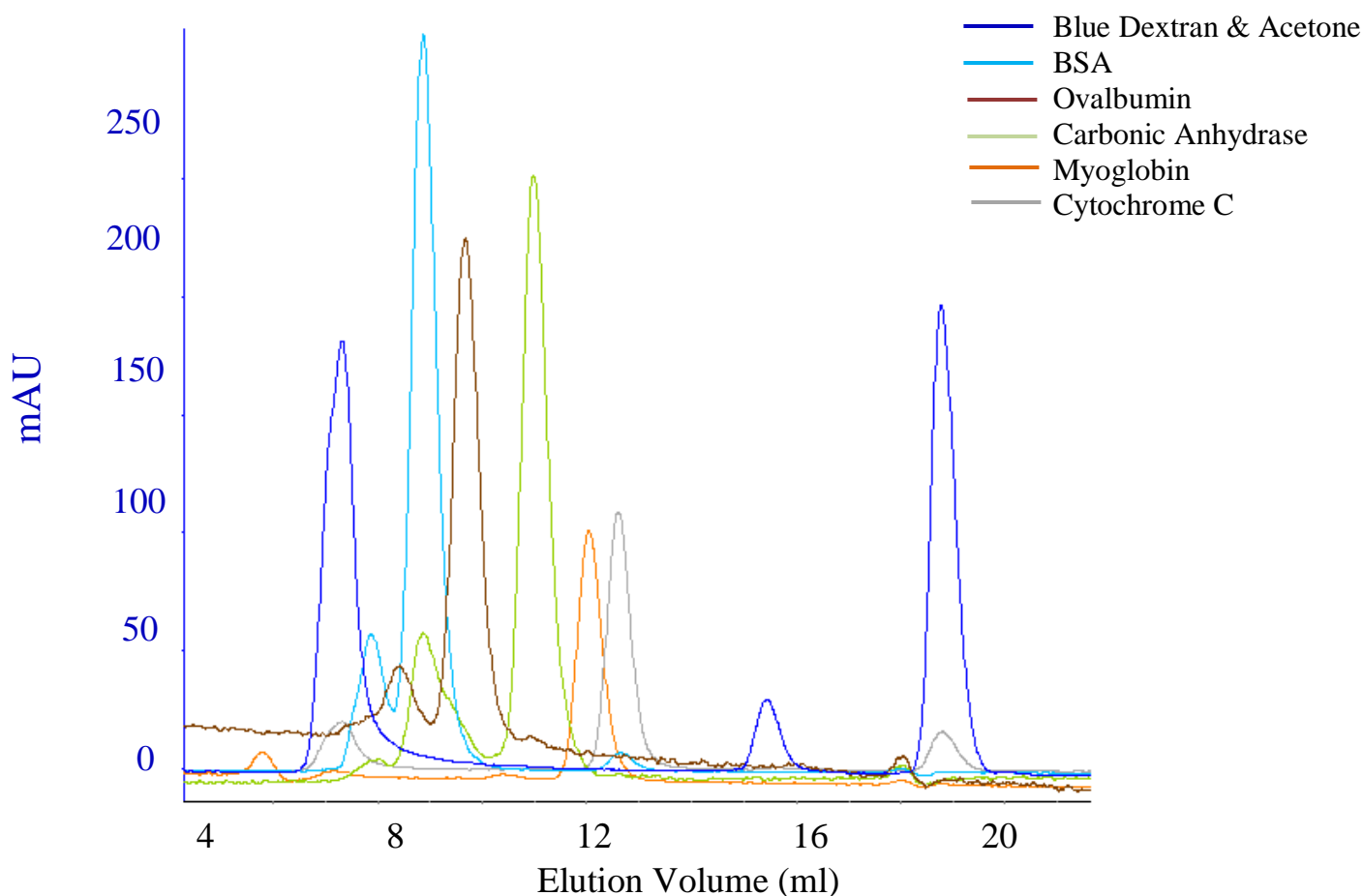


Fig 4.3 Elution of protein standards on Superdex 75 column. The elution profiles of standard proteins using gel permeation chromatography with the Superdex 75 column. The optical density at OD_{280nm} of eluted protein is plotted against the elution volume. The elution peak volumes were used in the construction of the standard curve (Fig 4.2)

4.4 Investigation of lectin compatibility with Superdex 75 GPC column

The Superdex 75 10/300 GL gel permeation chromatography column (GE Healthcare; Uppsala, Sweden) is composed of cross-linked agarose and dextran (Fig 4.4), oligosaccharides for which a lectin could display an affinity. To determine if binding was occurring between a lectin and the column matrix, PA-IL (Fig 4.6) and PL-IL (Fig 4.7) were applied to the column under normal conditions in the presence of competing raffinose sugar for which both lectins have a high affinity (Section 5.2). The sample and mobile phase both contain 50 mM raffinose to ensure competition throughout the entire run. Any difference in elution volume between the sugar replete and sugar deplete samples could then be accounted for by matrix-lectin interaction.

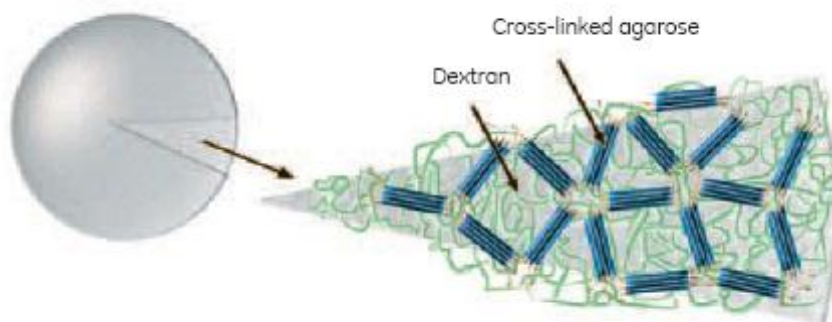


Fig 4.4 Schematic view of a superdex 75 bead composed of dextran and cross linked agarose. Average bead size is 13 μ m. (Image taken from Amersham Ltd.).

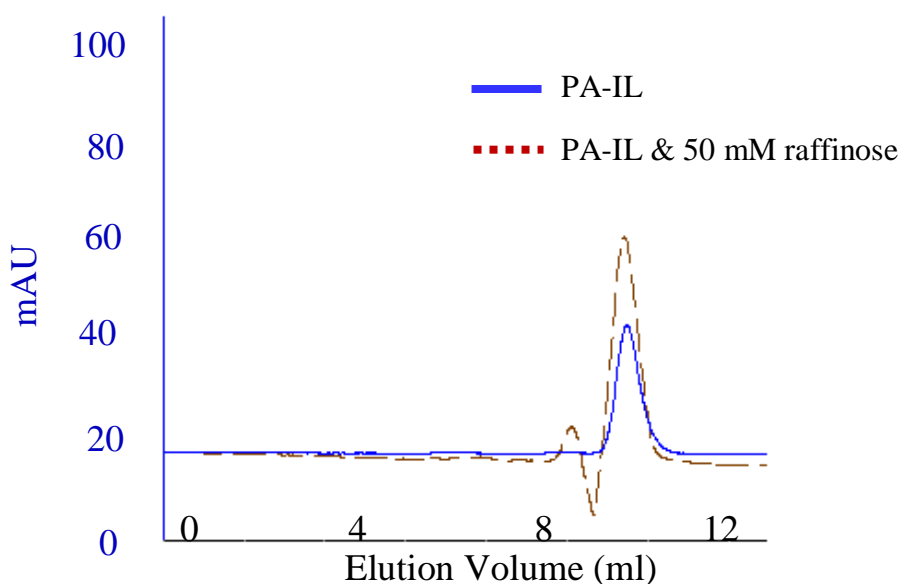


Fig 4.5 The adsorption $A_{280\text{nm}}$ of PA-IL through the Superdex 75 10/300 column in the presence/absence of raffinose. The elution profile of PA-IL30 (Section 3.3.1) is indicated with a brown line, with the profile of the same lectin in 50mM raffinose highlighted in blue. Peak maxima are at 11.425 ml and 11.487 ml respectively.

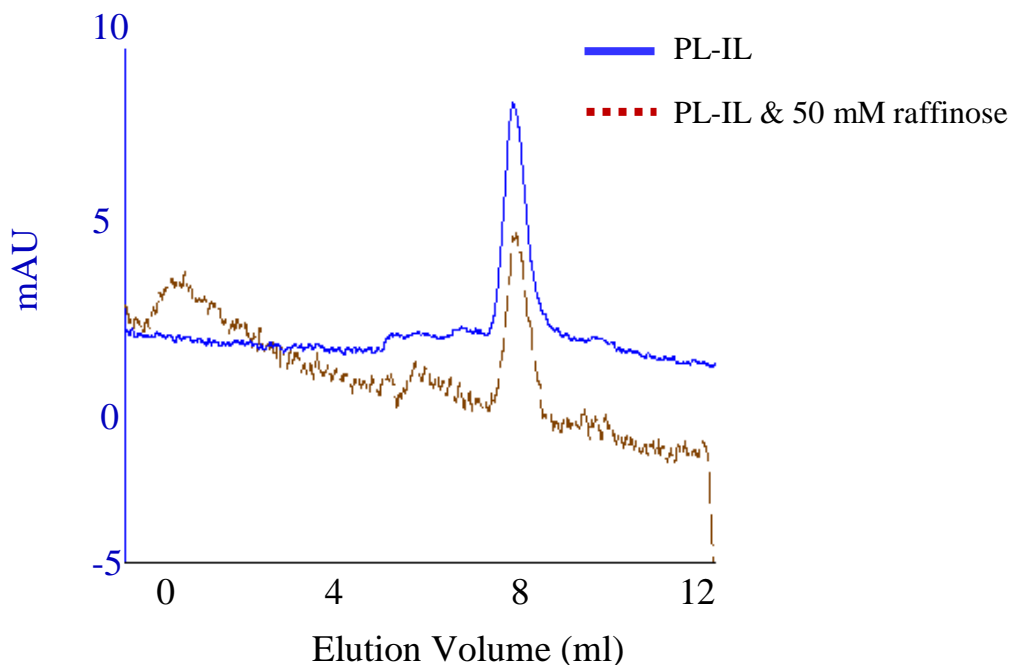


Fig 4.6 The adsorption $A_{280\text{nm}}$ of PL-IL through the Superdex 75 10/300 column in the presence of raffinose. The elution profile of PL-IL30 (Section 3.3.1) is indicated with a brown line, with the profile of the same lectin in 50mM raffinose highlighted in blue. Peak maxima are at 9.882ml and 9.829 ml respectively.

It was found that both PA-IL had the same elution volume in raffinose replete and raffinose deplete conditions in the Superdex 75 column (Fig 4.5). Similarly, the PL-IL lectin did not elute any sooner in the presence of raffinose (Fig 4.6). This indicates no sugar-specific interaction with the column matrix which would have otherwise contributed to false elution volume values later in the study.

4.5 Determination of lectin size using Superdex 75 GPC

As outlined in Section 2.27, 100µl of 1 mg/ml lectin was analysed using gel permeation chromatography. The lectin was suspended in PBS, pH 7.2 in order to ascertain its multimeric structure under native conditions. Having elucidated that no sugar-specific interactions were occurring between the lectins and the gel matrix, evaluation of lectin size was undertaken. Samples were applied to the column in 100 µl volumes at a concentration of 1 mg/ml, in various PBS-based buffers (Section 2.20), and the peak heights recorded were converted into relative molecular weights using the standard curve created in Section 4.2.

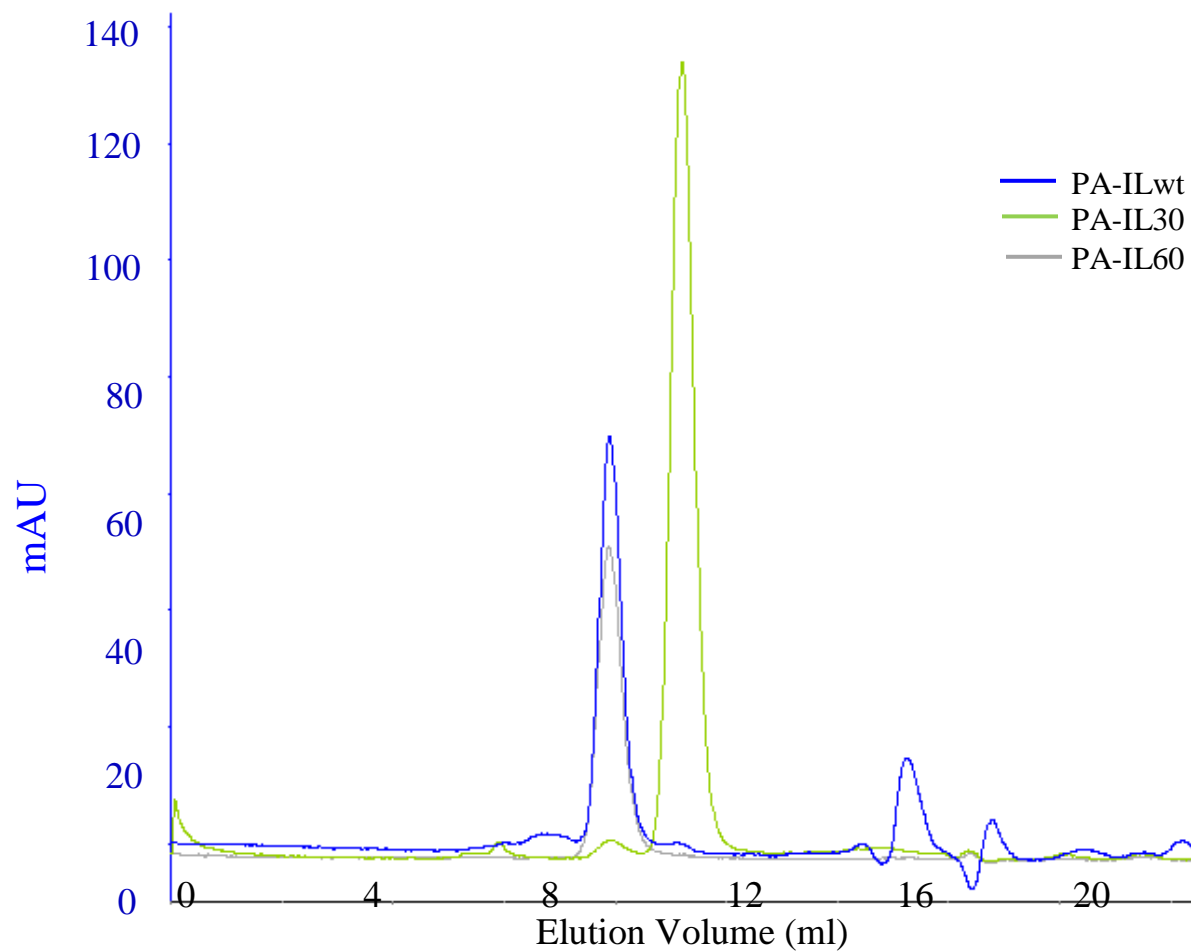


Fig 4.7 Elution Profile of PA-IL. The OD online analysis of purified PA-IL in PBS pH 7.2, passing through the Superdex 75 10/300 gel filtration column. Three large peaks can be seen. In blue is the elution profile of PA-ILwt with an elution volume of 9.87 ml, in grey the elution profile of PA-IL60 with an elution volume of 9.91 ml, and in green the elution profile of PA-IL30 with an elution volume of 11.48 ml. Impurities are visible as smaller peaks at 17 ml elution volume.

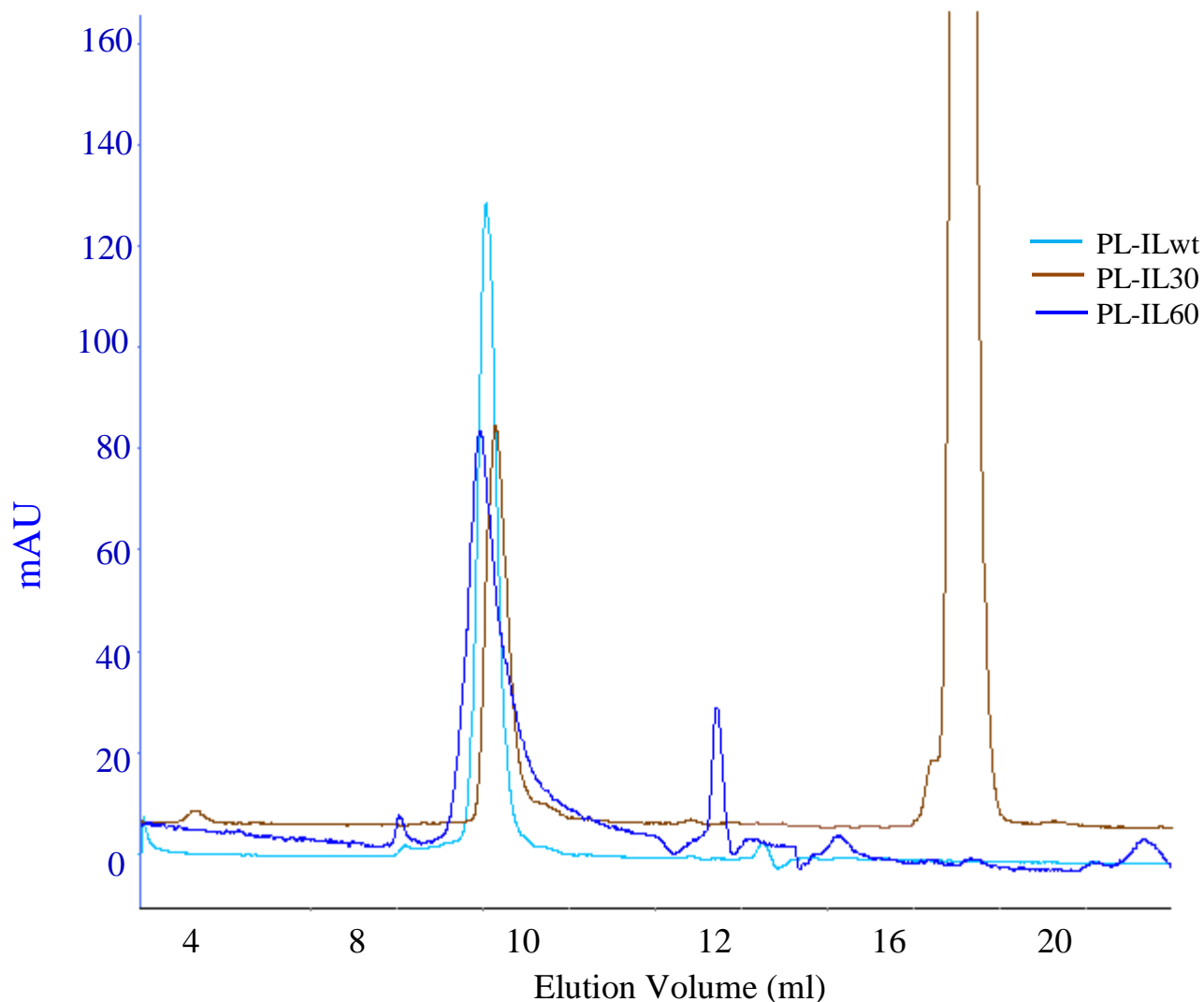


Fig 4.8 Elution Profile of PL-IL. The elution profile of recombinant PL-IL using gel permeation chromatography with the Superdex 75 column. The optical density at OD_{280nm} of eluted protein is plotted against the elution volume. Three large peaks can be seen. In cyan is the elution profile of PL-ILwt with an elution volume of 9.78 ml, in dark blue, the elution profile of PL-IL60 with an elution volume of 9.75 ml with a smaller peak at 12.3 ml, and in brown, the elution profile of PA-IL30 with an elution volume of 10.03 ml. A large acetone peak can be seen at 19 ml. Acetone is added to some samples as an internal control.

Table 4.2 Average Molecular Weights Obtained from Superdex 75 GF Column

Lectin	Average Retention Time (ml)	Theoretical Monomer Size (Da)	Calculated Molecular Weight
PA-ILwt (Untagged)	9.87	12761.8	44418.1
PA-IL60 (C-Tagged)	9.68 (± 0.03)	13828.3	41884.2
PA-IL30 (N-Tagged)	11.484 (± 0.05)	14291.9	20518.5
PL-ILwt (Untagged)	9.78	12884.5	44657.8
PL-IL60 (C-tagged)	9.75 (± 0.02)	13881.5	45321.9
PL-IL30 (N-Tagged)	10.03 (± 0.16)	14356.0	39941.04

The calculated molecular weights for each recombinant lectin are provided in Table 4.2. The molecular weights were calculated by measuring the average elution volume at which the maximum peak height was eluted from three separate runs and correlating these with the standard curve created in Section 4.3. Each lectin was run on the column separately. Their elution profiles are plotted together in Fig 4.7 and Fig 4.8. These molecular weights are not clear multimers of the actual monomer sizes (for example the predicted tetramer size for PA-ILwt is 51 kDa with the trimeric species at 38 kDa with the measured size of the protein was 44 kDa). All molecules were determined to have relative molecular masses below the predicted size of a tetramer with the exception of PA-IL30, which is measured half-way between the expected size of a monomer and a dimer. The differences between the predicted molecular weights of each monomer are a result of the (His)₆ tag and of spacer residues that exist between the tag and the lectin. The pQE30 vector incorporates a MRGS sequence before the *N*-terminal (His)₆ tag and a GS sequence between the lectin and tag, while the pQE60 plasmid encodes the RS residues upstream of the *C*-terminal (His)₆ tag).

4.6 Validation on Superdex GF results using a different column matrix.

Although it was shown that the lectins PA-IL and PL-IL did not interact with the matrix in the Superdex 75 gel filtration column (Section 4.5), the values obtained for the relative molecular masses of each protein were not comparable to predicted relative molecular masses based on their amino acid sequences (Table 4.2). It was decided to use another gel filtration matrix to examine protein size and to investigate if any non-specific interactions were causing the unexplained protein sizes.

The Toyopearl HW-50S (Tosoh Biosciences, Germany) column matrix is composed of methyl-methacrylate (the compounds structure is displayed in Fig 4.9) , and has no carbohydrate component. As a result, it could be said that this matrix would be more suitable for the examination of carbohydrate binding molecules.

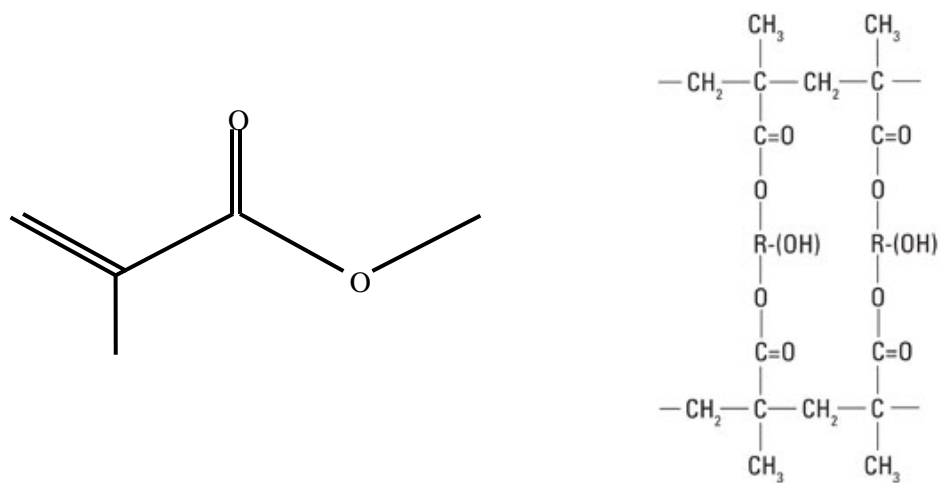
The HW-50S column had to be poured manually and as a result an extra column validation step had to be carried out. Column performance was checked at regular intervals, as manually poured columns can be susceptible to compaction over time. The number of theoretical plates (N), the figure used to validate column efficiency, was determined using the following equation.

Equation 4.2

$$N = 5.54 \left(\frac{V_e}{V_{1/2}} \right)^2 \times \frac{1000}{L}$$

In this equation, V_e corresponds to the retention volume; $V_{1/2}$ is the peak width at half the peak height; and L the peak height (mm). Depending on the matrix type, typical values for column performance will vary. Values are calculated by running 0.5 ml of 0.1% (v/v) acetone over the column at the flowrate to be used for running samples. The retention time measured is at OD_{280nm}. The resulting elution profile can be seen in Fig 4.10. Values for L, V_e and $V_{1/2}$ are measured from this profile and imputed into equation 4.2, to obtain N. This is then compared to manufacturers operators parameters to determine if the column is operational.

Having validated the column, a standard curve was again created using the method outlined in Section 4.3 with the same set of standard proteins. The elution profile for each standard can be seen in Fig 4.11 with the values for each given in Table 4.3



Note: R = Hydroxylated Aliphatic Group

Fig 4.9 Monomeric and polymeric structure of methyl methacrylate. Stick diagram of the monomeric methyl methacrylate and the hydroxylated polymeric version that forms the matrix in the gel filtration chromatography column of HW50s.

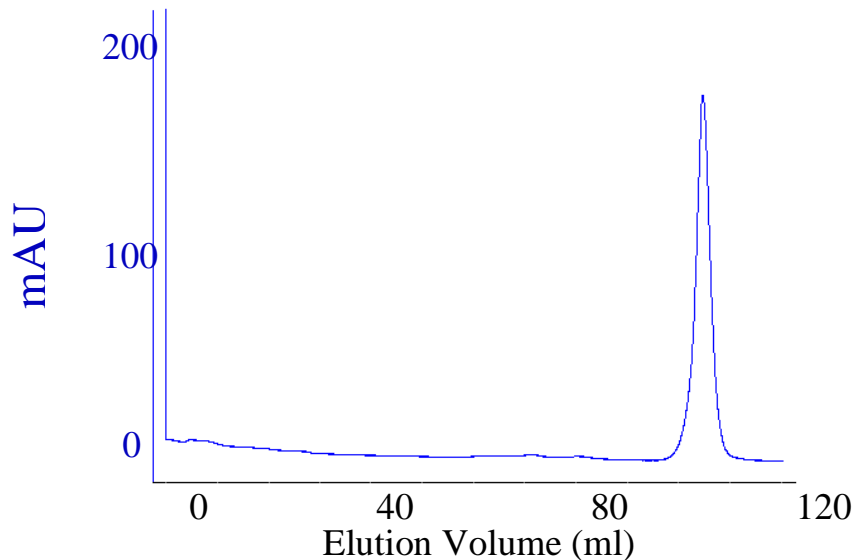


Fig 4.10 Determination of column efficiency. The elution profile of acetone (0.1% w/v) using gel permeation chromatography with the Toyopearl HW50 column. The optical density at OD_{280nm} of eluted acetone is plotted against the elution volume. Void volume (elution volume for acetone) was determined to be 104.82 ml.

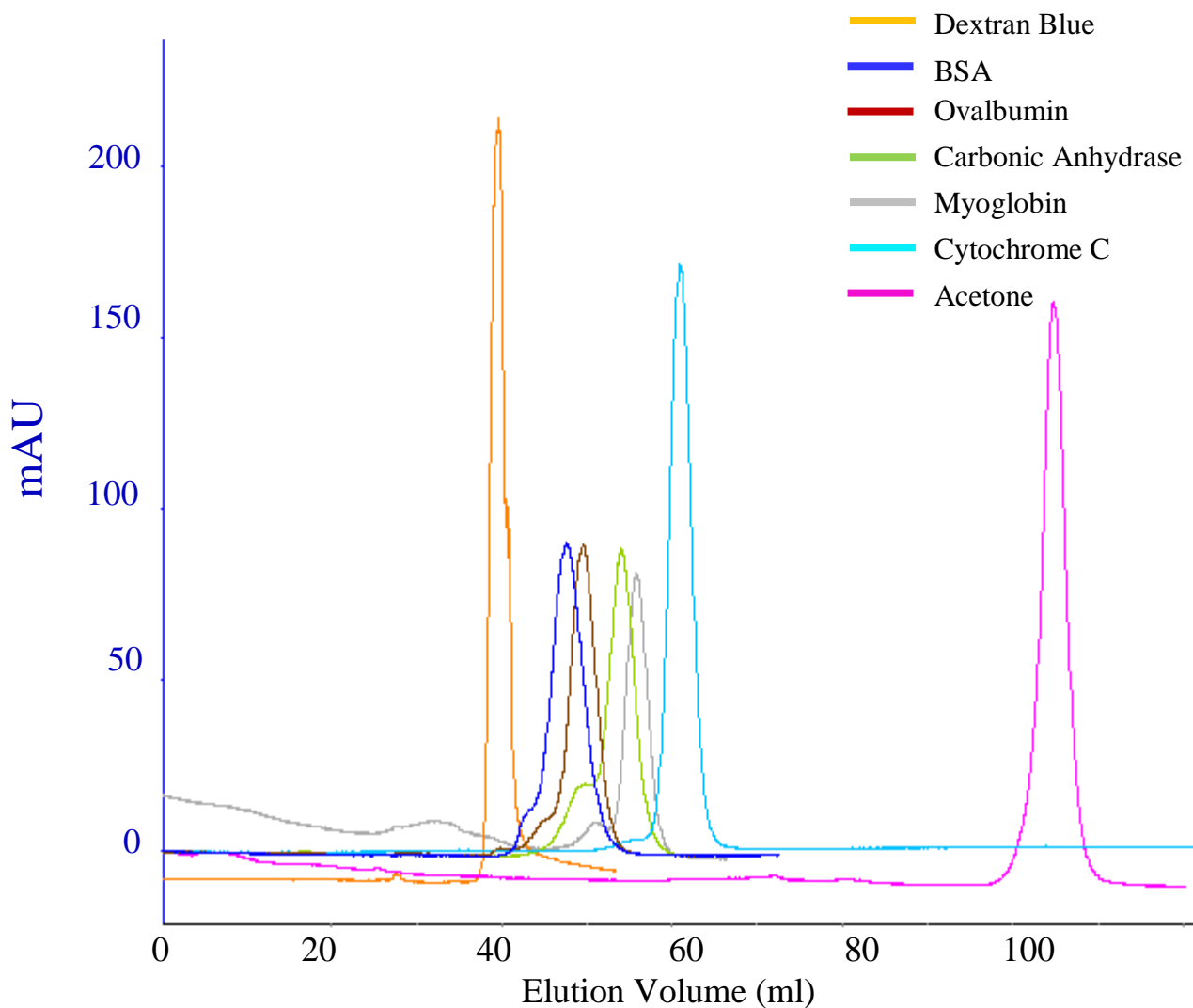


Fig 4.11 Elution of Protein Standards from the Toyopearl HW-50S GPC column. The elution profiles of standard proteins using gel permeation chromatography with the Toyopearl HW50 column. The optical density at OD_{280nm} of eluted standard proteins are plotted against the elution volume. The standard proteins are the same as those used in Section 4.4, The elution volume for each protein was used in the construction of the standard curve (Fig 4.12).

Table 4.3 Construction of a protein molecular weight standard curve for the Toyopearl HW50S gel permeation chromatography column at pH 7.2 in PBS

Standard	Elution Volume	Expected MW	logMW	K _a	Actual MW	%Error
Cytochrome C	60.88 ml	12400	9.43	0.67	12376	-0.19
Myoglobin	55.87 ml	17600	9.78	0.75	21978	-0.41
Carbonic anhydrase	54.12 ml	29000	10.28	0.78	28881	14.92
Chicken albumin	49.47 ml	45000	10.71	0.84	51717	24.24
BSA	47.54 ml	66000	11.10	0.87	65848	-0.23

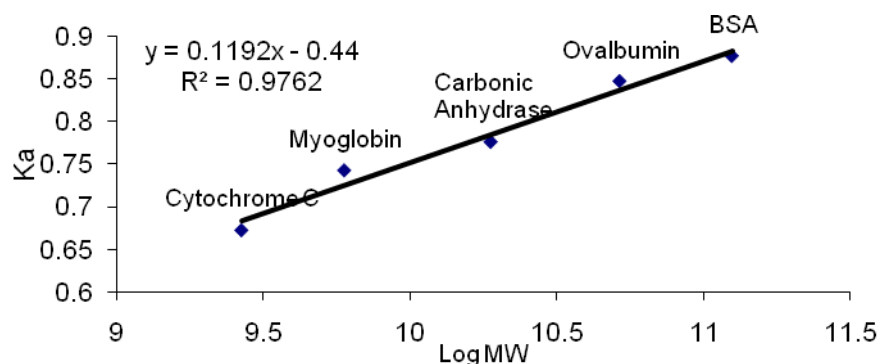


Fig 4.12 Development of size exclusion chromatography standard curve for the Toyopearl HW50S GPC column. The protein molecular weight standard curve was created using size exclusion chromatography, with the Toyopearl HW50S GPC column. Void volume (V_o) of 39 ml was determined from the elution volume of blue dextran, and fill volume (V_t) was 105 ml, determined with acetone. The elution volumes of protein standards are given in Table 4.3. These values were used to construct a plot of K_a versus log MW. This could be in turn used to calculate the relative MW of all lectin samples that were eluted in this range.

Prior to any investigations using the Toyopearl HW GPC column, its efficiency needed to be determined as the column had been poured manually. It is not uncommon for air-pockets or ineffective packing to occur during this process. Using equation 4.2 and the values obtained from Fig 4.10, the theoretical plate count for the resin was determined to be 2014, which was within the operable range for this resin. Before the relative molecular mass of any unknown protein could be evaluated, a molecular weight standard curve was required. This was constructed as before (Section 4.3). The R^2 value for the standard curve was 0.97. The standards used were in the range of the expected size of monomeric and tetrameric lectins.

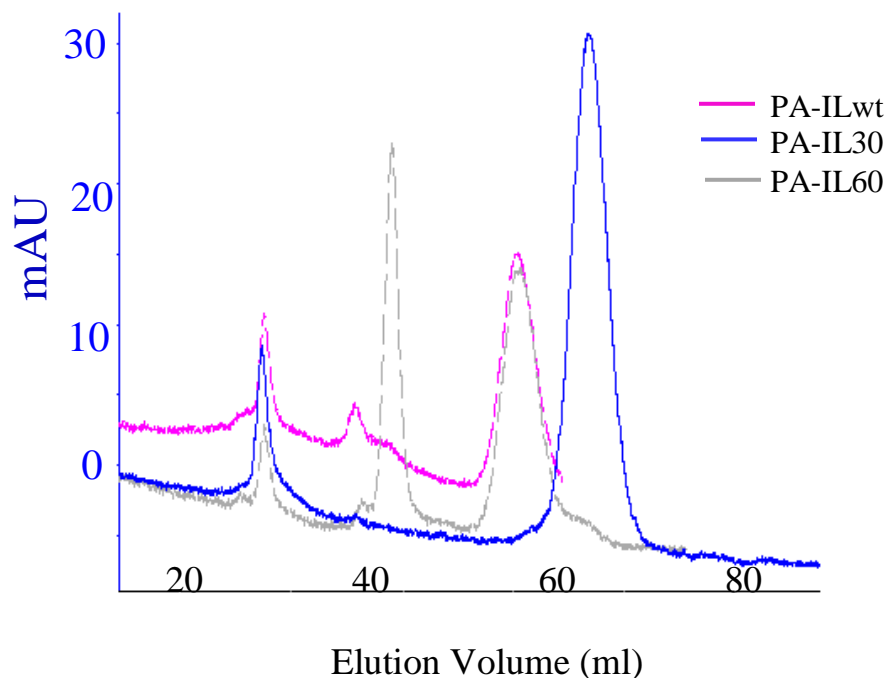


Fig 4.13 Elution profile of PA-IL in 0.2 M PBS pH 7.2 (Toyopearl HW-50S column matrix). The elution profiles of PA-IL proteins using gel permeation chromatography with the Toyopearl HW50 column. The optical density at OD_{280nm} of the eluted proteins are plotted against the elution volume. The positioning of the (His)₆ tag at the N-terminus has a drastic effect on the tertiary structure, as the main peak for the recombinant lectin is shown to be significantly smaller than the C-terminally tagged and untagged counterparts. Also identifiable here are varying forms of high molecular weight multimers that elute earlier in the gel filtration which elute outside the range of the standard curve..

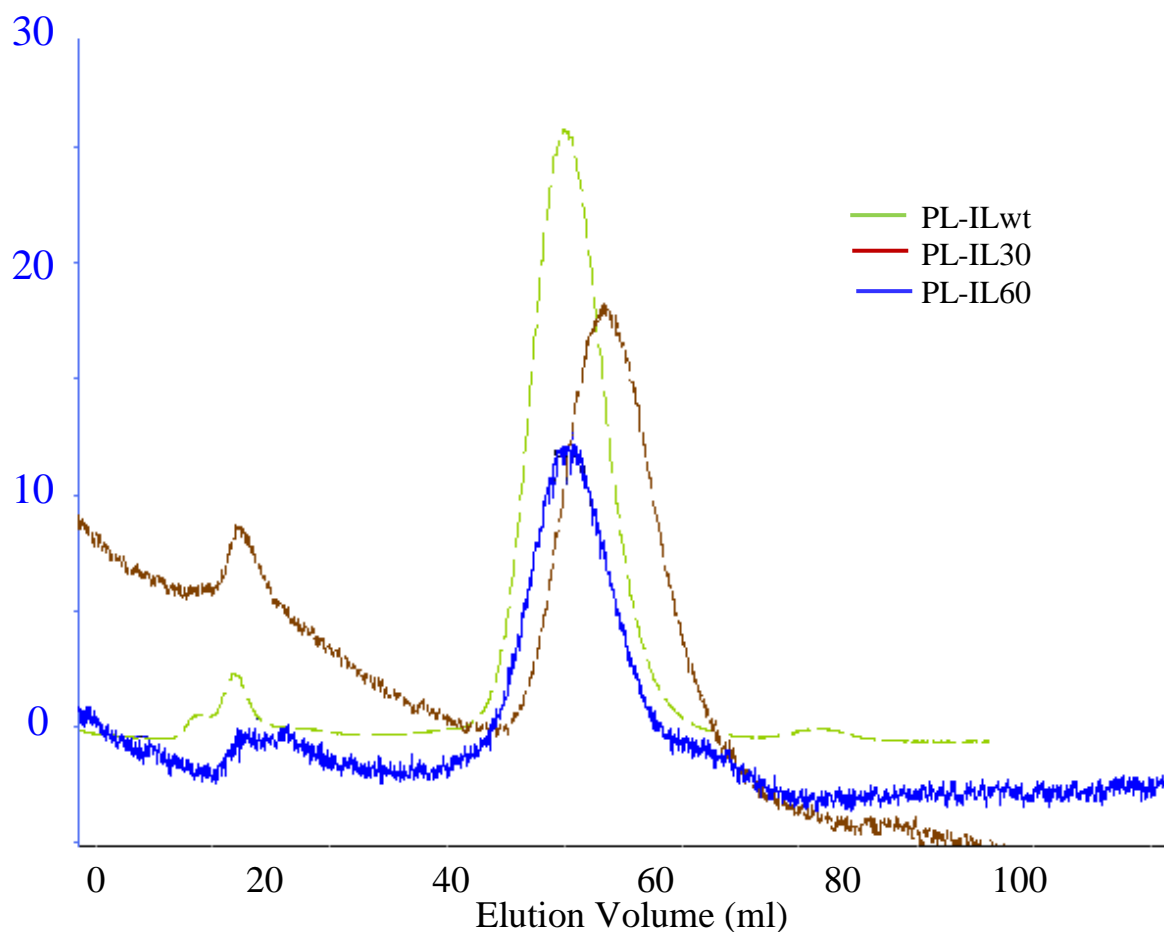


Fig 4.14: OD 280nm elution profile of PL-IL in 0.2M PBS pH 7.2. (Toyopearl HW-50S column matrix). The elution profiles of PL-IL proteins using gel permeation chromatography with the Toyopearl HW50 column. The optical density at OD_{280nm} of the eluted proteins are plotted against the elution volume. In Blue, C-terminally tagged PL-IL, in brown N-terminally tagged PL-IL, and in green untagged PL-IL. The three lectin variants can be seen to elute in close proximity suggesting the position of the recombinant (His)₆ tag had little effect.

Table 4.4: Lectin sizes calculated using Toyopearl HW-50S column

Lectin	Average Retention Time (ml)	Theoretical Monomer Size (Da)	Calculated Molecular Weight
PA-IL Untagged	50.15 (± 0.0)	12761.8	47493.9
PA-IL C-Tagged	50.525 (± 0.1)	13828.3	44749.9
PA-IL N-Tagged	57.265 (± 0.3)	14291.9	19476.0
PL-IL Untagged	49.86 (± 0.2)	12884.5	50818.3
PL-IL C-tagged	49.96 (± 0.1)	13881.5	49809.7
PL-IL N-Tagged	52.18 (± 1.3)	14356.0	45536.1

Having successfully created a standard curve for the Toyopearl GPC column (Fig 4.12), the relative molecular masses of the different recombinant lectins could be estimated. PA-ILwt, PA-IL30 and PA-IL60 were run separately three times with three representative elution profiles seen in Fig 4.13. The average elution volumes for each form of PA-IL lectin are given in Table 4.4 with the respective relative molecular masses that correspond to an elution volume also provided. In agreement with the Superdex 75 results (Section 4.5), the PA-ILwt and PA-IL60 molecules were found to form high molecular weight molecules smaller than the expected size of a tetramer (51 kDa and 55 kDa) while the PA-IL30 molecule forms a 20 kDa molecule which falls between the expected size of a monomer (14 kDa) and a dimer (28 kDa). The PL-IL recombinant molecules respective elution profiles can be seen in Fig 4.14 and the result obtained from each provided in Table 4.4. The PL-IL results also correspond with those seen on the Superdex 75 column as they form structures that lie between the expected sizes of trimers and tetramers.

4.7 Determination of PA-IL_{mut1} quaternary structure by GPC

In the PA-IL_{mut1} the amino acids predicted to play a role in carbohydrate binding in PL-IL were introduced into PA-IL. Before investigating the binding properties of the mutant lectin it was important to compare the quaternary structures of the wild-type with the mutated PA-IL molecules. The proteins were analysed by GPC using the HW50S column and their respective elution profiles are presented in Fig 4.15.

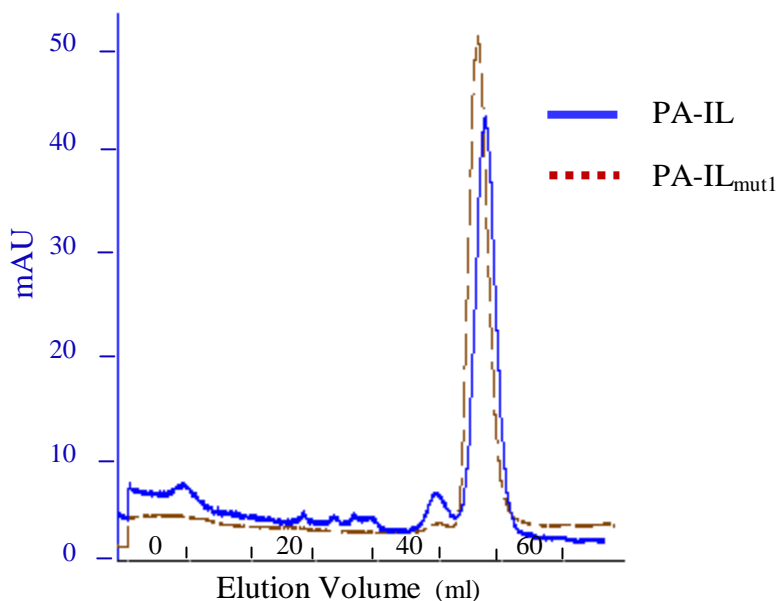


Fig 4.15 Comparison of PA-IL₃₀ and PA-IL_{mut1}₃₀ elution profiles. The elution profiles of PA-IL₃₀ and PA-IL_{mut1}₃₀ using gel permeation chromatography with the Toyopearl HW50 column. The optical densities at OD_{280nm} of the eluted proteins are plotted against the elution volume. The elution volume of PA-IL₃₀ was 57.31 and the elution volume PA-IL_{mut1}₃₀ was 56.98. Due to the amino acid changes introduced by mutagenesis, the mutant should be 78.2 Da larger per monomer.

PA-IL₃₀ and PA-IL_{mut1}₃₀ showed no significant difference when both were examined on the HW-50S column matrix. According to the standard curve created in Section 4.6, the relative molecular masses of PA-IL_{mut1}₃₀ was 20.1 kDa compared to 19.5 kDa determined for PA-IL₃₀. Consequently, it could be concluded that the three amino acids changes introduced in the mutant were not effecting the formation of quaternary structures in the molecule. This result means the binding affinities and specificities of both molecules can be directly compared as they have the same valencies. This is not the case when tetrameric and dimeric molecules are compared.

4.8 Determination of lectin quaternary structure by mass spectrometry

Mass spectrometry is a method used to determine the relative masses of atoms or molecules. An electrical charge is placed on the molecule and the resulting ions are separated by their mass to charge ratio (m/z). In Section 3.3.2 mass spectrometry was used to evaluate the precise mass of purified proteins by examining the charge/mass ratio. Lectins are known to form non-covalent oligomeric complexes in their native state and thus the assembly of the oligomeric complex needs to be quantified in order to determine the number of subunits present in the molecular assembly.

Careful optimization of operating parameters was required to observe intact non-covalently associated molecular assemblies as the integrity of the molecular assemblies needs to be preserved during phase transition from the solution to the gas phase.

The sample source is one such parameter and a standard source and a nano-flow source were compared. The nanoflow source was preferred as it provides desolvation (the release of the water electrostatically bound to a particle in solution) of ions from aqueous solvents at room temperature. Sensitivity is also higher due to higher desolvation efficiency as the flow volumes are much lower. An infusion rate of 200 nL/min and a capillary voltage of 2.75 kV were found to be optimal for this application. The standard source proved to have broader peaks, resulting from the presence of solvent adducts due to ineffective desolvation and the inherently low resolving power of the TOF analyzers at higher m/z values. A sample analysis of PA-IL using the standard source is shown in Fig 4.16.

PA-IL is known to exist in its native state as a tetramer and so 'native' conditions were optimized for it. It was found that all forms of PA-IL and PL-IL could then be examined under the same electro-spray (ES) conditions and in the same solvent (formic acid 0.1% (v/v)) without sample dissociation.

A mass spectrometry method named collision-induced dissociation (CID) was used to break up ions in the gas phase. It dissociates the ions, in this case charged

proteins, by accelerating them under high electric potential in the vacuum of the mass spectrometer, before collision with a neutral gas. The transfer of kinetic energy to internal energy within the analyte results in bond breakage. By using increments of higher collision energies, quaternary structure disruption is brought about and the degree of multi-valency can then be accurately estimated. Thus the relative internal strength of a molecule can be estimated. Through changing the collision energy imparted on the lectin molecule, it was then possible to analyze the extent of oligomerisation of PA-ILwt, PA-IL30, PA-IL60, PL-ILwt, PL-IL30 and PL-IL60. Results obtained could then be compared with those obtained from the GPC described in Sections 4.5 and 4.6.

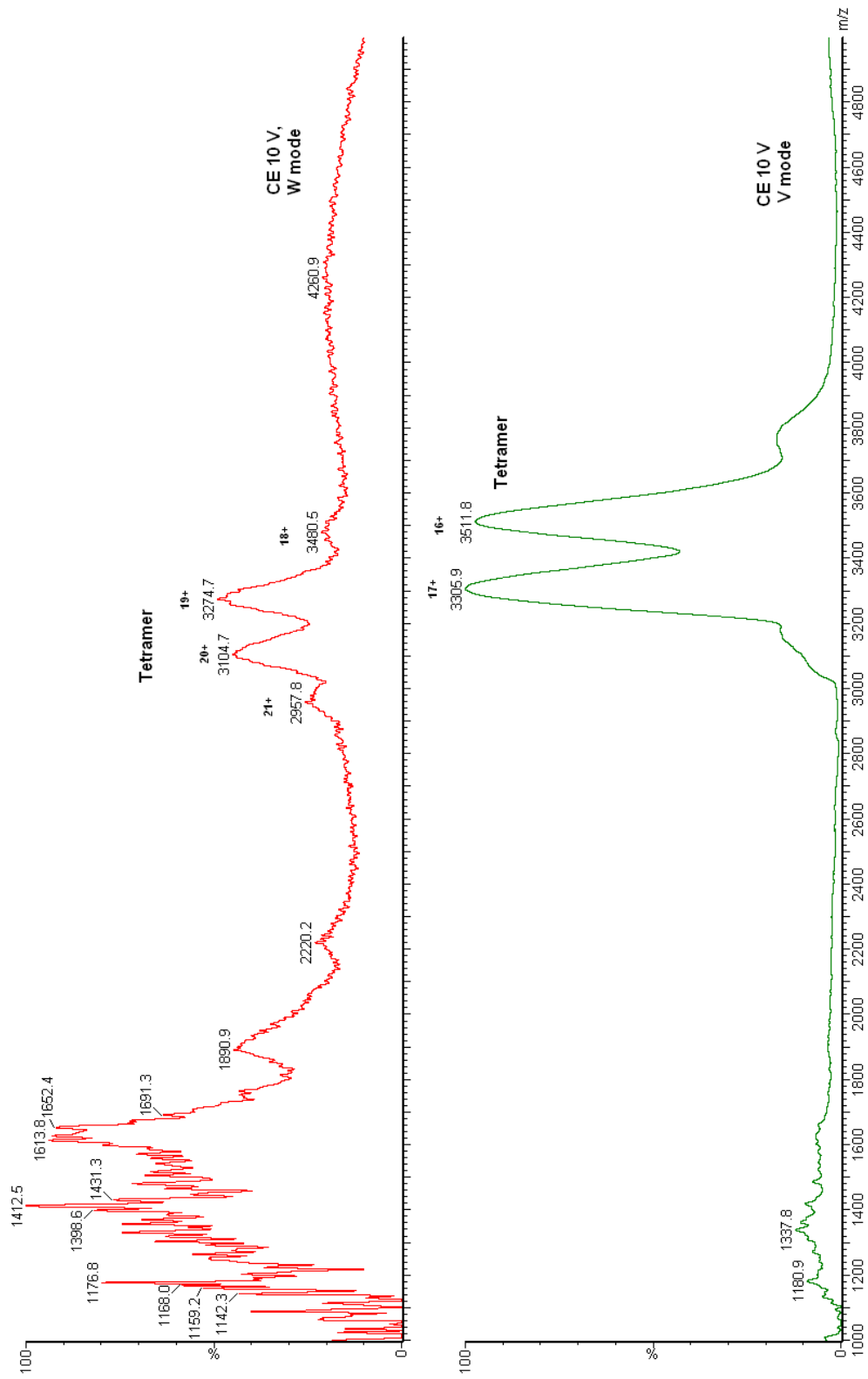


Fig 4.16: Mass spectra of PA-ILwt in ammonium bicarbonate (pH 7) in W and V modes, using standard source at Pi 5.2 mbar. PA-ILwt was analyzed using the standard ESI interface and ammonium bicarbonate buffer (pH 7.0) as a solvent to mimic “native like” conditions. (V and W modes refer to the shape of the ion path within the spectrometer).

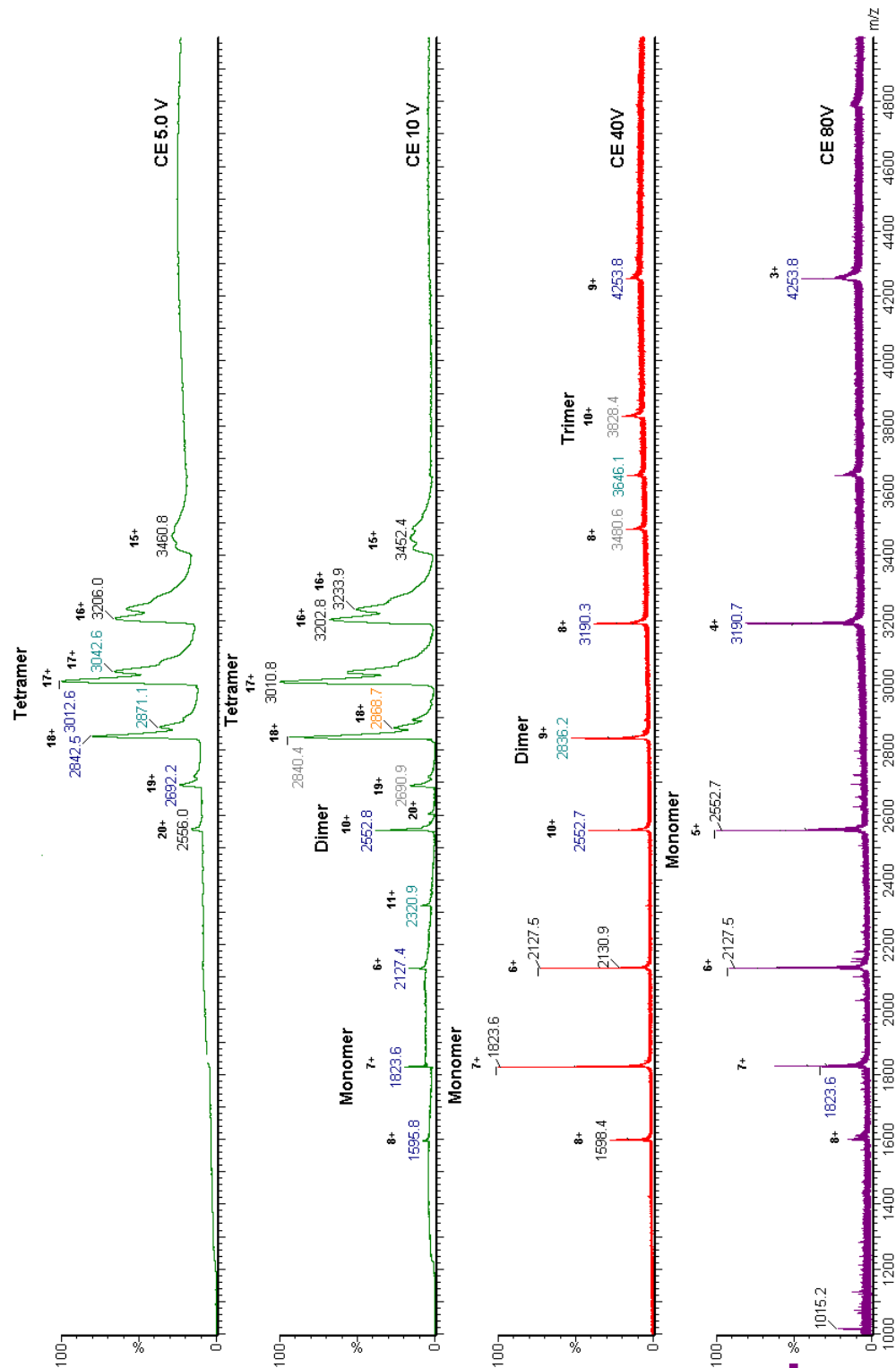


Figure 4.17: Dissociation of PA-IL tetramer ions into trimer, dimer and monomer ions by CID. Image depicting the break-up of the PA-IL-wt molecule at various different collision energies. Individual peaks such as that seen at 2127 (which was calculated to have the molecular weight corresponding to a monomer) increase in size as the amount of monomer in the sample increases, while the tetramer peak at 3012 disappears as the tetrameric molecule is disrupted.

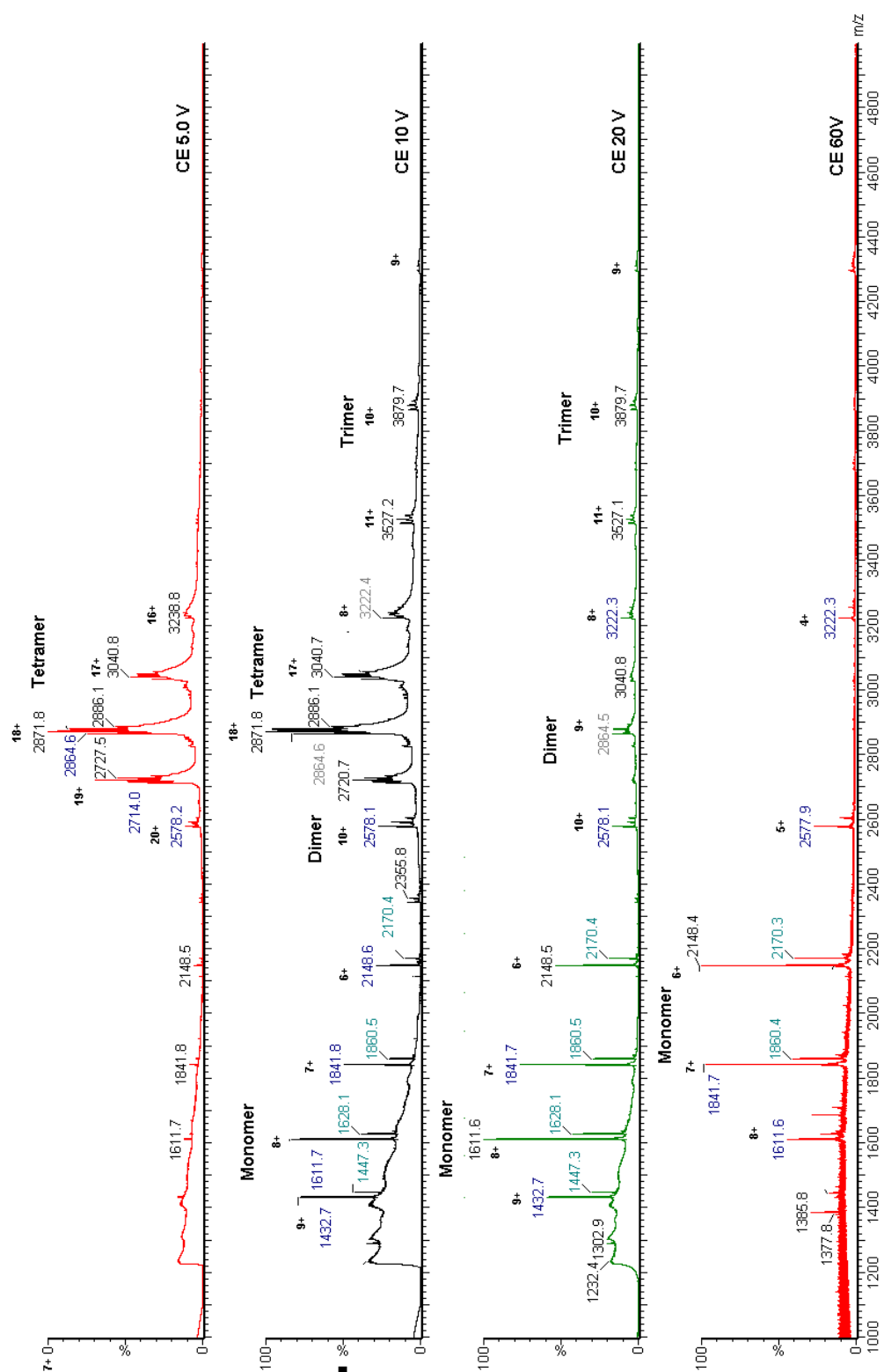


Fig 4.18: Dissociation of PL-IL60 tetramer ions into trimer, dimer, and monomer ions by CID. Image depicting the break-up of the PL-IL60 molecule at various different collision energies. Individual peaks such as that seen at 2148 (which was calculated to have the molecular weight corresponding to a monomer) increase in size as the amount of monomer in the sample increases, while the tetramer peak at 3040 disappears as the tetrameric molecule is disrupted.

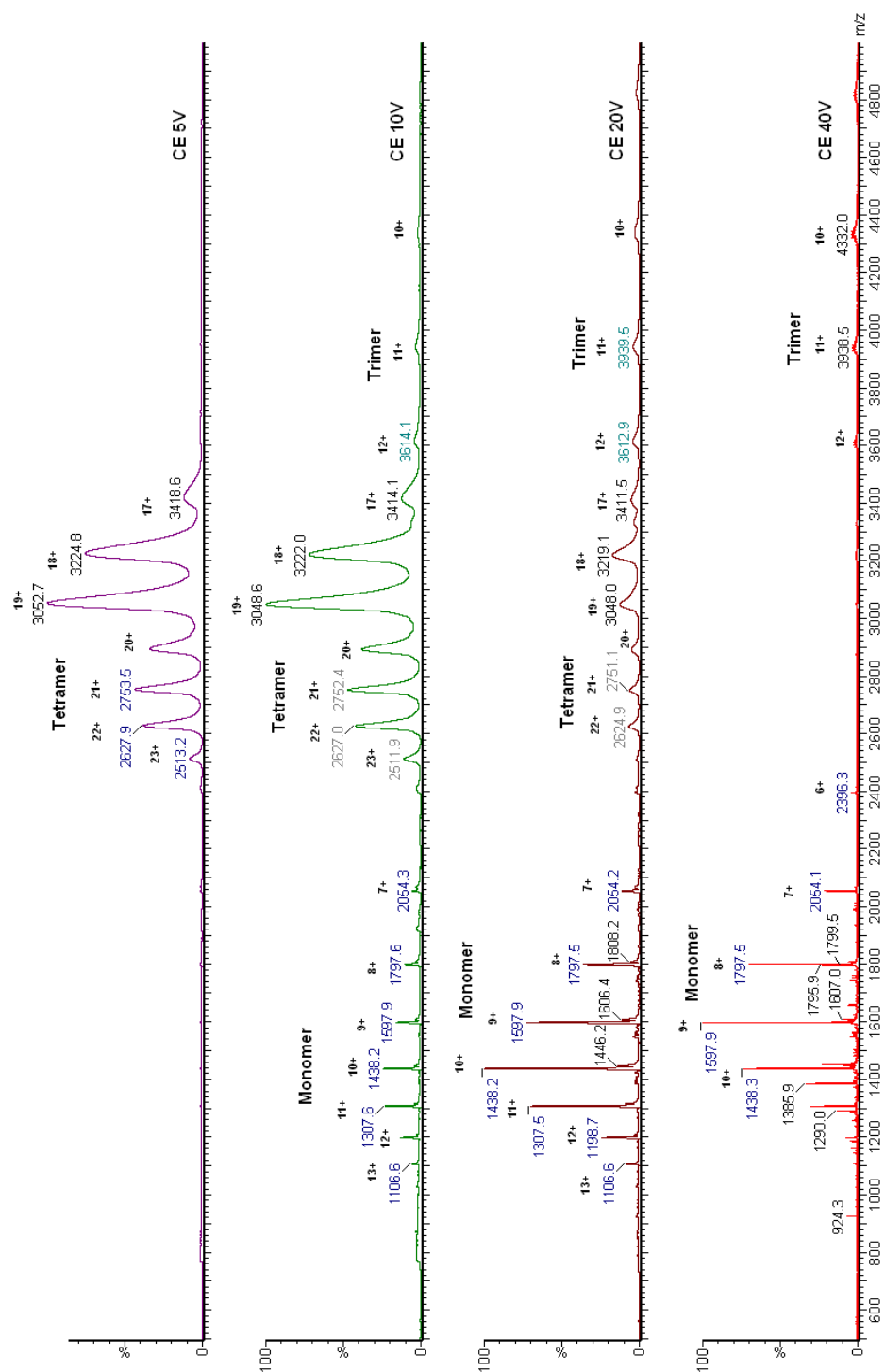


Fig 4.19: Dissociation of the PL-IL30 tetramer ions into trimer and monomer ions by increasing collision voltage. Image depicting the break-up of the PL-IL30 molecule at various different collision energies. Individual peaks such as that seen at 1597 (which was calculated to have the molecular weight corresponding to a monomer) increase in size as the amount of monomer in the sample increases, while the tetramer peak at 3224 disappears as the tetrameric molecule is disrupted.

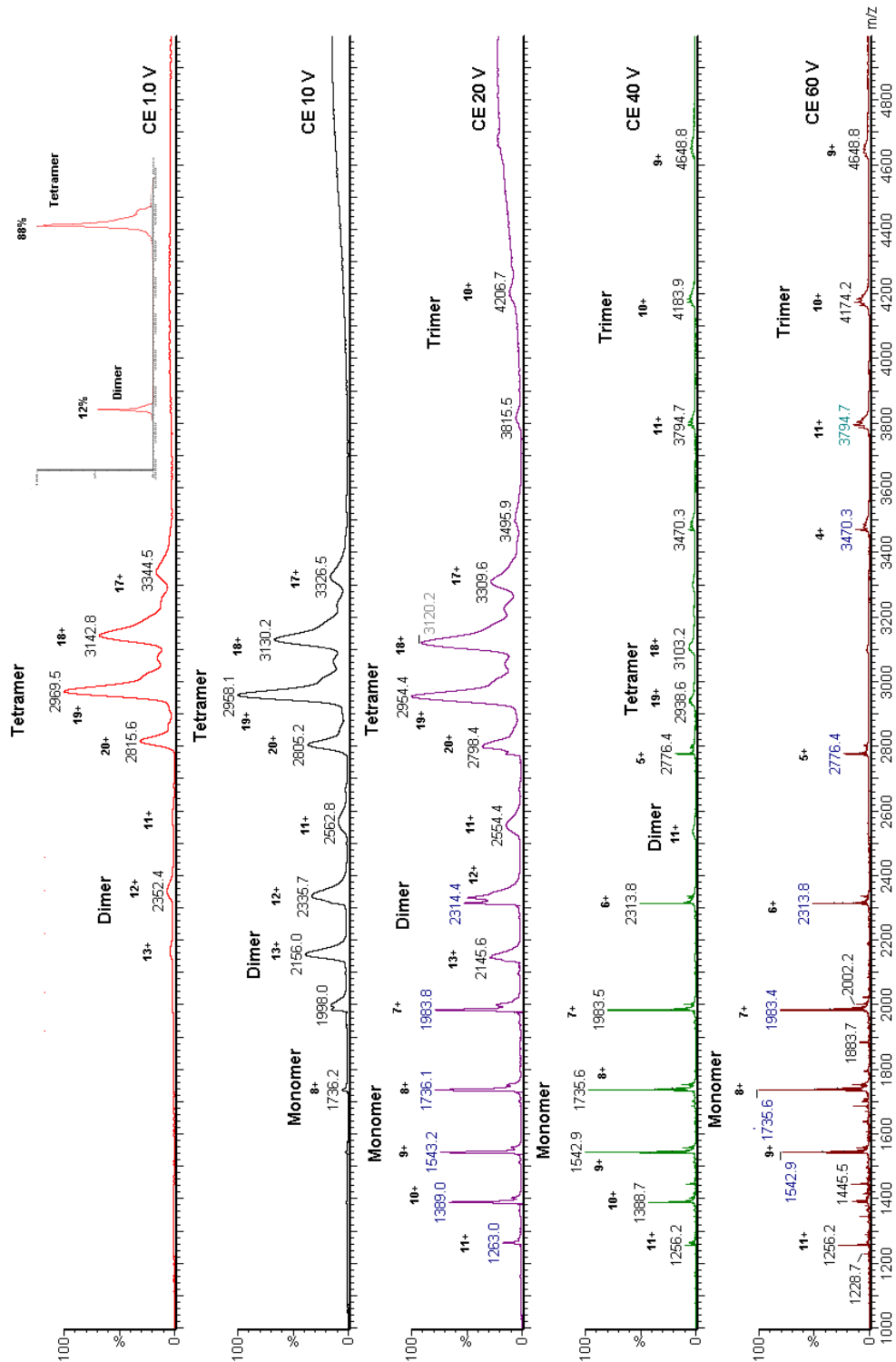


Table 4.5: Summary of lectin quaternary structures observed by mass spectrometry

Lectin	Molecular assembly observed	Dissociation voltage (V)	Products observed upon dissociation
PA-IL	Tetramer (100%)	10 V	Trimer, Dimer, Monomer
PA-IL60	Dimer (100%)	10 V	Monomer
PA-IL30	Dimer (100%)	5 V	Monomer
PL-IL	Tetramer (100%)	5 V	Trimer, Dimer, monomer
PL-IL60	Tetramer (88%)	5 V	Trimer, Dimer, Monomer
	Dimer (12%)	10 V	Monomer
PL-IL30	Tetramer (100%)	10 V	Trimer, Monomer

Fig 4.17 shows that at sufficiently low accelerating potential (< 5 V) mass spectra of PA-ILwt deconvoluted to 51139.14 ± 0.0 Da. It was later shown that commercially available wild-type PA-IL deconvoluted to 51665.1 ± 6.5 (methionine is cleaved in PA-ILwt which results in the apparent mass difference). At higher accelerating potential (10 V) the tetrameric species was partially dissociated into the corresponding dimeric and monomeric species. At 40 V, the tetrameric species were completely dissociated into the trimeric species at 38270.5 ± 10.9 Da, the dimeric species at 25516.6 ± 0.4 Da and the monomeric species at 12758.7 ± 0.8 Da. Under the same operating conditions as for PA-ILwt, the mass spectrum for PA-IL60 displayed only a dimeric form (Table 4.5). The dimer ion could withstand collision activation up to 5 V and started to dissociate at 10 V into a monomeric form. Further collisional activation at 40 V produced lower charge state product ions corresponding to the monomeric species at 13828.3 ± 1.5 Da (calc. av. mass 13828 Da) and the dimeric species (27801.8 ± 40.05 Da).

Similarly to PA-IL60, even at low accelerating potential (1 V), the mass spectrum of PA-IL30 displayed the presence of only dimeric (94%) and monomeric ion (6%) species, with the tetrameric species being completely absent (Table 4.5). The dimeric form started dissociating at 5 V into the monomeric form while at 40 V, dimer ions were completely dissociated into the monomer ions. The higher charge

state monomer ions started to fragment at this potential. Under ‘solvent stripping’ conditions monomer and dimer ions deconvoluted to 14308 ± 0.4 (calc. av. mass 14292) and 28874.8 ± 3.4 Da, respectively. The observed mass for the monomer is 16 Da higher than the calculated mass, consistent with retention and subsequent oxidation of the terminal methionine residue to methionine sulfoxide in the purified lectin protein. This has been reported in other *E. coli* derived proteins (Spector *et al.*, 2003).

The dissociation pattern of PL-ILwt is similar to PA-ILwt, wherein the tetramer ion dissociates into the trimer, dimer and monomer ions with asymmetric distribution of charge (Fig 4.20). However, the collision activation required to dissociate the PL-IL tetramer ion was appreciably less (5 V vs 10 V) when compared to PA-IL, suggesting that PA-ILwt is more stable compared to PL-ILwt.

PL-IL60 exists as both a tetramer (88%) and a dimer (12%) in its native state (Fig 4.18). At increasing collision energies (5 V and 10 V) the tetramer began to dissociate into dimers and monomers, with only monomer visible at CE 40 V. In contrast, PL-IL30 retained its structural integrity until at a collision energy of 20V. The trimeric species were still detectable at CE 40V, suggesting it is a slightly more stable protein its C-tagged counterpart (Fig 4.19).

A comparison of the results obtained by ES-MS with those obtained by GPC is given in Table 4.6. Assuming that the GPC determines the actual sizes of tetramers to fall below what would be expected, the results obtained between the two methods agree, with the exception of PA-IL60. This result is examined further in Section 4.9.

Table 4.6 Comparison of ES-MS results with those observed by GPC

Lectin	ES-MS	Superdex 75	HW-50S
PA-ILwt	Tetramer	Tetramer	Tetramer
PA-IL30	Dimer	Dimer	Dimer
PA-IL60	Dimer	Tetramer	Tetramer
PL-ILwt	Tetramer	Tetramer	Tetramer
PL-IL30	Tetramer	Tetramer	Tetramer
PL-IL60	Tetramer & Dimer	Tetramer	Tetramer
PA-IL _{mut1} 30	Dimer	Dimer	Dimer

4.9 Disruption of quaternary structure by pH and salt concentration

The quaternary structures determined by gel filtration in Section 4.4 and Section 4.5 and by mass spectrometry observed structures in Section 4.6 shows discrepancies only in terms of the molecule PA-IL60. Each other assembly showed matching structures for all techniques. Therefore, the PA-IL60 molecule was investigated under more expansive conditions.

The GPC experiments to date had been conducted at pH 7.2 in PBS, relatively neutral conditions when compared to the conditions required for ionisation in the mass spectrometer, which required 0.1% (v/v) formic acid. The PA-IL60 molecule was then investigated by GPC, under acidic as well as high salt concentration solutions, in order to evaluate if the harsher conditions could account for the break up of the tetrameric molecule seen in the mass spectrometer (Fig 4.22), and the formation of the dimeric molecule seen in mass spectrometer analysis (Section 4.8). The exact conditions used in ES-MS could not be repeated in GPC, as the operating parameters for the HW-50s and Superdex 75 columns do not allow for pH's ranging below 3. Similarly, the buffering capacity of PBS buffers is at its lowest point at pH 5.6. Therefore, citrate buffer at pH 4 was used to examine the effect of pH on the molecule. For comparison, the apparently more stable PA-ILwt molecule was also studied under the same conditions (Fig 4.21).

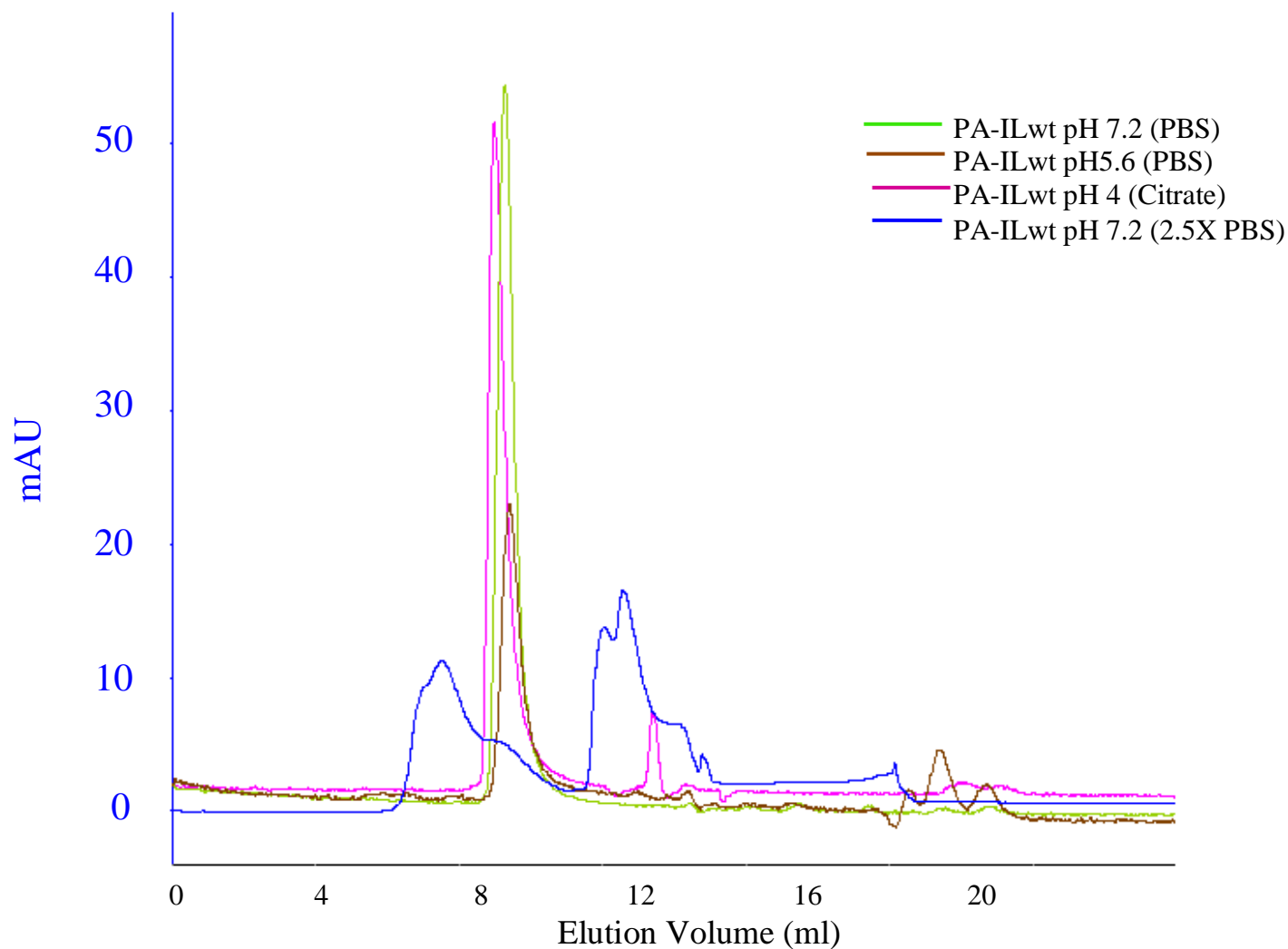


Fig 4.21: Elution volumes of PA-ILwt under varying conditions. The elution profiles of PA-ILwt using gel permeation chromatography with the Superdex 75 column using various buffer conditions. The elution volume of PA-ILwt was unaffected by dropping the pH to 5.6, and only slightly by dropping to pH 4, with severe break visible under high salt concentration,

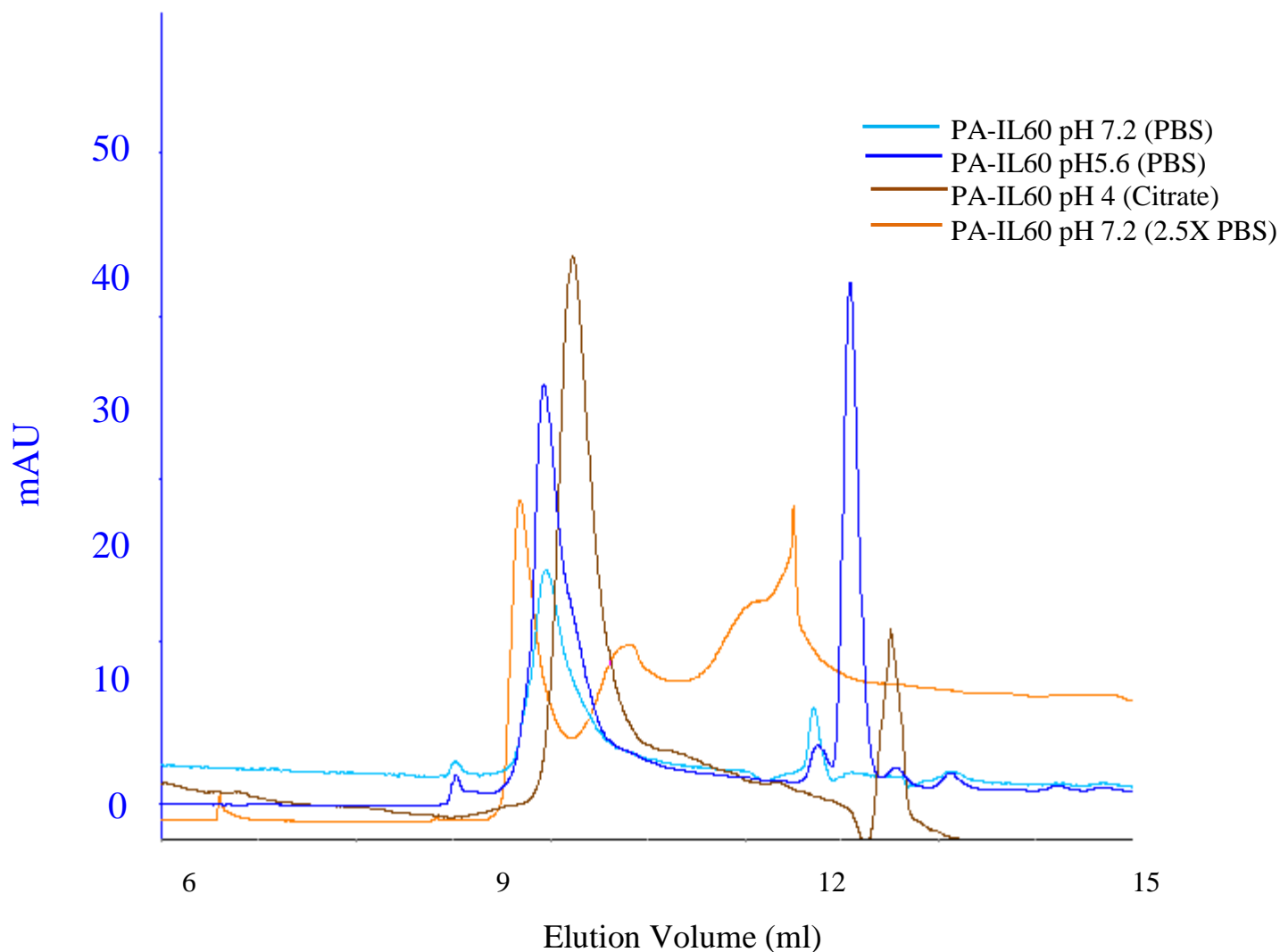


Fig 4.22 Elution volumes of PA-IL60 under varying pH and salt conditions. The elution profiles of PA-IL60 using gel permeation chromatography with the Superdex 75 GPC column under various buffer conditions. PA-IL60 has one elution peak at 9.9 ml. When the PBS pH is dropped to 5.6, a second peak is visible at 12 ml, the size of a monomer. When exposed to citrate buffer pH 4, the tetrameric peak moves to 10.6 ml, with the monomer peak shifting by the same amount. However, with increased salt concentration there was a dramatic break up with these peaks not being as well defined.

4.10 Determination of the valency of lectins according to the hemagglutination assay

Lectins form structures consisting of two or more identical or highly similar subunits, each containing one carbohydrate binding site with the same specificity (Sharon and Lis, 2002). It is this property that allows many lectins to agglutinate erythrocytes, with each subunit (protomer) binding to one of the array of glycans coating the erythrocyte. Thus, the quantity of lectin required to agglutinate a fixed concentration of RBCs will be directly proportional to its valency. Traditionally, this assay was used to characterise the sugar specificity of a lectin by the addition of competing free sugars. However, we propose that the assay can also be used to provide information about the type of oligomer assembly present in a lectin molecule. In Table 4.6, the relative molecular masses for each of the recombinant lectin forms determined by previous methods were compared. To calculate the valency of the lectins forms, hemagglutination was utilised (Section 2.23). A fresh 100 µg/ml of each lectin stock was used for each study and serially diluted until the hemagglutination limit was reached. Table 4.7 shows the minimum quantity for each lectin form required to agglutinate 50 µl of 3.5% (v/v) RBCs.

Table 4.7 Quantities of lectin required for hemagglutination

Lectin	Minimum Concentration Required for Hemagglutination	Quaternary structure estimated from GF and MS
PA-ILwt	0.4 µg (±0.06 µg)	Tetramer
PA-IL60	0.6 µg (±0.1 µg)	Tetramer
PA-IL30	0.1 µg (±0.05 µg)	Dimer
PL-ILwt	1.2 µg (±1.0 µg)	Tetramer
PL-1L30	0.9 µg (±0.2 µg)	Tetramer
PL-1L60	0.8µg (±0.15 µg)	Tetramer

4.11: Discussion:

Lectins are often describes as Velcro™ type molecules. Although the lectin-carbohydrate interaction often has a weak affinity this is compensated for by their ability to form multivalent molecules that can bind tightly to the targets. Consequently, any disruption to a lectins quaternary structure can have serious effects on the molecules binding properties and hence its potential on analytical platforms. For functional analysis of these lectins it is important to have an understanding of the degree of multi-valency of the molecule to be investigated. This knowledge can greatly improve the understanding of their functional analysis and also aid in the comparison of different lectins.

Initial gel filtration studies were carried out using the carbohydrate matrix Superdex 75. After confirming that lectins PA-IL and PL-IL eluted similarly from the column in the presence and absence of raffinose (Fig 4.5 and Fig 4.6) it was concluded that carbohydrate-lectin interactions were not impeding protein migration through the column.

It was then shown, in the case of PL-IL, that the location of the (His)₆ affinity tag did not hinder the formation of the native tetramer structure as seen in the untagged version (Fig 4.8). This was not the case for PA-IL, where an N-terminal (His)₆ tag had completely impeded the ability of the lectin to form structures higher than a dimer (Fig 4.7).

As can be seen from Table 4.2, the actual relative molecular masses are consistently smaller than those predicted from the amino acid sequence. The one exception being PL-IL60, which is only slightly larger. It is theorized that as the standard curve was created using globular proteins, it serves as only a rough estimate as to the size of the non-globular shaped lectins. The PA-IL molecule (Section 1.9.2) and from theoretical modeling, the PL-IL molecule (Fig 4.1) are not globular but instead form long narrow structures. As a result it is probable they do not migrate through the column matrix at the same rate as globular proteins. When the size of the PA-IL30 was examined, on both GPC matrices, a molecule of around 20,000 Da was estimated. As the monomer is 14 kDa, and a dimer 28 kDa, this value suggests a

truncated dimer of some type. However, if the *N*-terminally tagged molecule forms a shape not resembling a globular protein, the standard curve will not apply and only broad conclusions are possible.

The Superdex and HW-50S GPC matrices, although both composed of different materials and employed in columns of different dimensions, gave similar results for all lectins tested (Table 4.2 and Table 4.4). It is unlikely that the same non-specific interactions are occurring between the lectins and two different matrices. Therefore, it can be reasonably concluded that the migration pattern of the lectins through these columns are attributable to relative molecular mass and shape only.

Noticeable on the HW-50S traces are the presence of high molecular weight aggregates (Fig 4.13 and Fig 4.14). This may be due to fact that with the HW-50S column being much larger than the Superdex column, in length and width, a higher concentration of protein was applied. It was often observed that when working with these lectins at concentrations of 1 mg/ml or greater, that they would often fall out of solution. It is possible that this is what is being seen in these traces, as according to the standard curve, these peaks correspond to proteins larger than dextran blue. In determining the stoichiometry of the lectins inconclusive results were obtained. For example, it was not conclusive whether PL-IL30 formed a trimer or tetramer or if PA-IL30 formed a monomer or dimer. Therefore, another technique was employed to investigate lectin assembly.

Mass spectrometry is a useful technique for protein mass determination but its use in the evaluation of multivalent structures is an emerging field. The technique involves the transition of lectins from liquid to gaseous phase and the removal of the water molecules associated with the protein can result in unnatural interactions. For the visualization of the different charged molecules within the electrospray of the mass spectrometer it was necessary to suspend the lectins in 0.1% (v/v) formic acid. This was also an environment that could lead to results that were not representative of the lectin in its native state. However, the results that were obtained by mass spectrometry, largely agreed with the gel permeation results (Table 4.6) with the exception of PA-IL60, which clearly did not form structures of a higher order than a

dimer when examined by ES-MS but formed tetrameric/trimeric structures when examined by GPC.

The mass spectrometry data also provided interesting information about the strength of the various assemblies as different collision energies were required to break up the oligomers depending on the stability of the structure.

Many groups have reported a seemingly odd dissociation behavior for large multimeric protein complexes and have suggested conformational flexibility of the monomeric subunit in the complex is responsible for this phenomenon (Felitsyn *et al.*, 2001; Jurchen and Williams, 2003). In the case of PA-ILwt, initial removal of a large amount of charge results in the formation of the lower charge state trimer ion and, as such, it appears at higher m/z values than the parent tetramer ion (Fig 4.18). With sufficiently high collision activation (>40 V), the trimer ion undergoes two successive dissociation steps, each reducing the charge state of the product ions and finally resulting in a single envelope corresponding to a monomeric species at 80 V. The most abundant charge state shifted from 7+ to 5+ which deconvoluted to 12758.7 ± 0.8 Da. This is in close agreement with the average molecular mass of 12762 Da for the monomer calculated from the amino acid sequence.

The dissociation patterns of PA-IL30 and PA-IL60 could only be partly investigated as the native state of both molecules, according to the ES-MS was a dimer. These dimers were relatively stable with relatively high low collision energies (40 V) required for the complete dissociation of the molecules.

A final method was then investigated for the determination of lectin assembly. Hemagglutination has been historically used as a technique to evaluate if a lectin had the ability to form multivalent structures. In this study it was used to differentiate between dimeric and tetrameric forms of the same lectin. This is done on the basis that the length, width and breadth of a monomeric unit are 167 Å, 50 Å and 51 Å respectively (Cioci *et al.*, 2003) with the diameter of a red blood cell at ~6 µm. As a result it is unlikely that a tetramer will have the ability to attach to more than two different red blood cells, with two sub-units in excess (Fig 4.23). With this in mind, two lectin solutions with similar lectin concentrations, have the ability to agglutinate to very different levels according to what their quaternary structure is. From the

lectin hemagglutination levels seen in Table 4.7 it is clear that the PL-IL molecule agglutinates at the same level regardless of tag location confirming the results seen by GPC and ES-MS. PA-IL results show that roughly twice as much untagged and C-tagged PA-IL is required for hemagglutination than N-tagged PA-IL. This suggests that the results seen in GPC are more representative of the lectins quaternary structure than the results seen in ES-MS. The amount of PA-ILwt required for hemagglutination is comparable to that found in the literature (5 μ g) for previous PA-IL studies (Giboa-Garber and Sudakevitz., 1999) giving further confidence to the results displayed.

As a final attempt to explain the difference between GPC and ES-MS results, the PA-ILwt and PA-IL60 molecules were analyzed by GPC under various conditions. Fig 4.22, shows that the molecule is more susceptible to break up than the wild-type (Fig 4.21). Varying conditions did not produce any peaks in the region where a dimer would be expected to elute. Therefore, conditions seen within the MS were not reproducible. However, it can be seen that the molecule does dissociate under acidic conditions and hence this could possible be the reason for the discrepancy within Table 4.6.

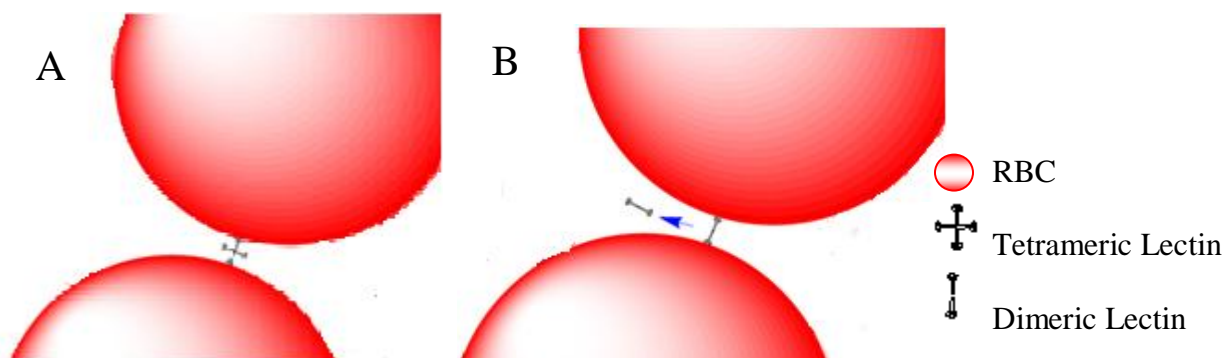


Fig 4.23 Illustration of hemagglutination by tetrameric and dimeric lectins. The size difference between the cell and the lectin are not to scale, but from the diagram it can be seen that in the hemagglutination reaction it is unlikely that one lectin molecule will be involved in the binding of more than two cells, due to steric hindrance. In the second diagram, it is shown how the dimeric molecule will agglutinate the red blood cells in the same manner, though if present in the same concentration as a tetrameric lectin, will be able to agglutinate twice as many red blood cells, due to there being twice as many 'active' molecules. Image is drawn using ChemBiodraw (CambridgeSoft Ltd)

5.0 Determination of the Sugar Specificity of Recombinant Lectins

5.1 Introduction

The cloning and expression of active recombinant PL-IL and PA-IL lectins is described in Chapter 3. Subsequent structural characterisation of these molecules was carried out through a number of complimentary methods (Chapter 4). The functional characterisation of these molecules is now explained through a series of specificity and affinity analysis experiments. It is important at this juncture to highlight that the specificity of a molecule refers to the selection of a ligand to its receptor, while the affinity relates to the strength of the attraction between ligand and receptor. Traditionally, lectins were characterised through the hemagglutination assay (Section 2.23). This method is utilised in this study, however the ELLA (Section 2.24) is proving a more useful method for glycobioologists, as it provides more detailed information about both lectin specificity and affinity.

5.2 Determination of lectin specificity by hemagglutination inhibition

The hemagglutination assay was introduced in Section 4.7, where it was utilised to investigate the valency of the recombinant lectins. The traditional use of the hemagglutination assay is for the determination of lectin specificity. It was using this technique that the first lectins were investigated and the human blood groups were identified in the 1950s (Section 1.6).

In summary, the lectins agglutinate red blood cells in a U-bottomed well (as opposed to red blood cells falling out of suspension by gravity), and the minimum amount of lectin required for the agglutination of a fixed quantity of red blood cells is termed one hemagglutination unit (HU). The fixed quantity of red blood cells used in this study was 3.5% (v/v), as some researchers use 3% (v/v) whilst others use 4% (v/v). The amount of each lectin required for 1 HU is already given in Table 4.5. A HU value for the mutant lectin PA-IL_{mut1} is not provided, as the quantity required for hemagglutination is in the region of 100 times greater than PA-IL or PL-IL, indicating a significant difference in affinity or specificity from PA-IL and PL-IL. It is worth noting at this juncture that the assay is dependant on the lectin having a high specificity for the sugars displayed on the red blood cell. Red blood cells from another animal, such as rabbit or mouse may display sugars which would have been more readily agglutinated by PA-IL_{mut1}, however rat red blood cells were the only erythrocytes available constantly. Consequently, PA-IL_{mut1} characterisation using the technique could not proceed. As seen later, other methods were thus employed.

Having determined the amount of lectin required for 1 HU (values displayed in Table 4.7), inhibition assays could be carried out. To directly compare the specificities of two lectins, it is vital that they have the same valency, as a result, untagged PA-IL and PL-IL were used for this study, as from Chapter 4, it was known that the untagged versions of each exist in the tetrameric form. The assay was carried out by the addition of 8 HU of lectin to a serial dilution of a free sugar, which was left to bind for twenty minutes before the addition of a fixed quantity of red blood cells. The higher the affinity of the sugar for the lectin, the less will be

required to inhibit the agglutination reaction. A primary screen of the main mono- and di-saccharides was carried out to investigate the specificity of the lectins (Table 5.2).

Having concluded that both lectins have a strong preference for galactose and glucose, more complex oligosaccharides and derivatised galactose molecules were employed in a more concentrated specificity study, the results of which are shown in (Fig 5.1). The structures of the more common oligosaccharides are provided in Table 5.1. Among the glycans tested, PL-IL reacted most strongly with were raffinose and stachyose as well as melibiose while PA-IL agglutination was most readily inhibited by the addition of monomeric galactoses with hydrophobic groups (phenyl- β Gal and nitrophenyl- β Gal).

It was also investigated whether the agglutination of red blood cells by PL-IL was in some way affected by variables such as temperature (Fig 5.2), as was found to be the case in a previous study, where PA-IL was found to be more active in the range 0°C to 20°C than at warmer temperatures (Gilboa-Garber and Sudakevitz, 1999). Due to the impracticalities of carrying out ELLA assays at colder temperatures, the results were not applied in terms of exploiting the higher activity, however this data is useful when discussing issues such as low activity and specificity later in the study.

Table 5.1 Structure of common oligosaccharides

Name	Structure	Name	Structure
Cellobiose	Glc β (1-4)Glc	Raffinose	Gal α (1-6)Glc β (1-2)Fru
Lactose	Gal β (1-4)Glc	Stachyose	Gal α (1-6)Gal α (1-6)Glc β (1-2)Fru
Maltose	Glc α (1-4)Glc	Melibiose	Gal α (1-6)Glc
Sucrose	Glc α (1-2)Fru		

Table 5.2 Hemagglutination inhibition against a range of sugars

Sugar	PA-IL	PL-1L
Agarose	X	X
Arabinose	X	X
Cellobiose	X	X
Fructose	X	X
Fucose	X	X
Galactose	√	√
GalNAc	√	√
GlcNAc	√	√
Glucose	√	√
Lactose	√	√
Maltose	√	√
Mannose	X	X
Ribose	X	X
Sucrose	X	X
Xylose	X	X

(√ denotes the ability to inhibit RBC agglutination at 5 mM concentration. X denotes no inhibition at the same concentration)

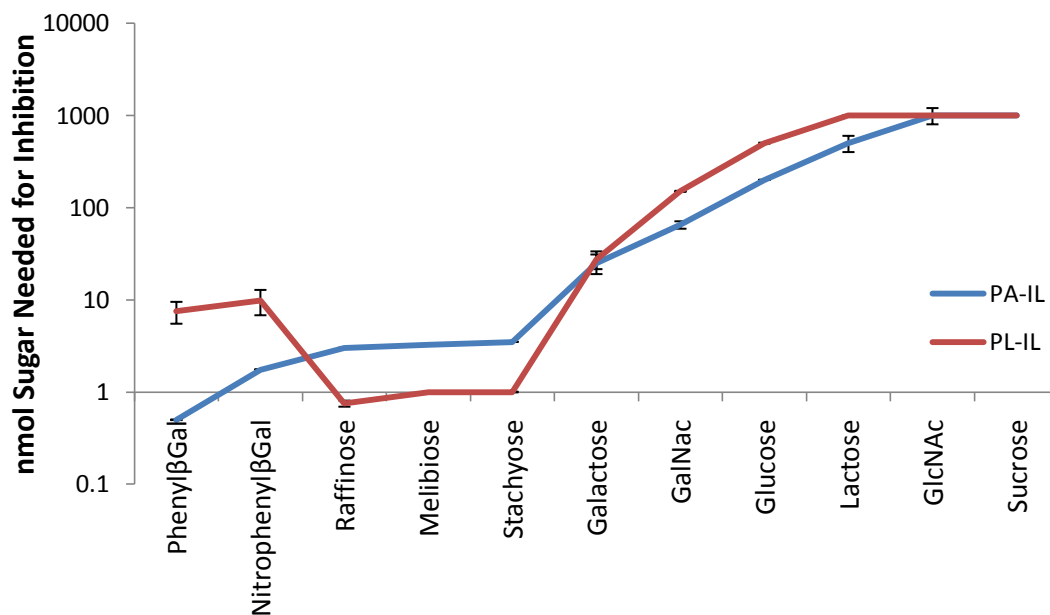


Fig 5.1 Hemagglutination inhibition of PA-IL and PL-IL using simple glycans. Graph showing the quantities of sugar (nmol) required to inhibit hemagglutination of the molecules PA-IL and PL-IL. Each lectin was used at a concentration of 8HU, to which serial dilutions were added. Inhibition is observed as red blood cells having fallen to the centre of the well through gravity.

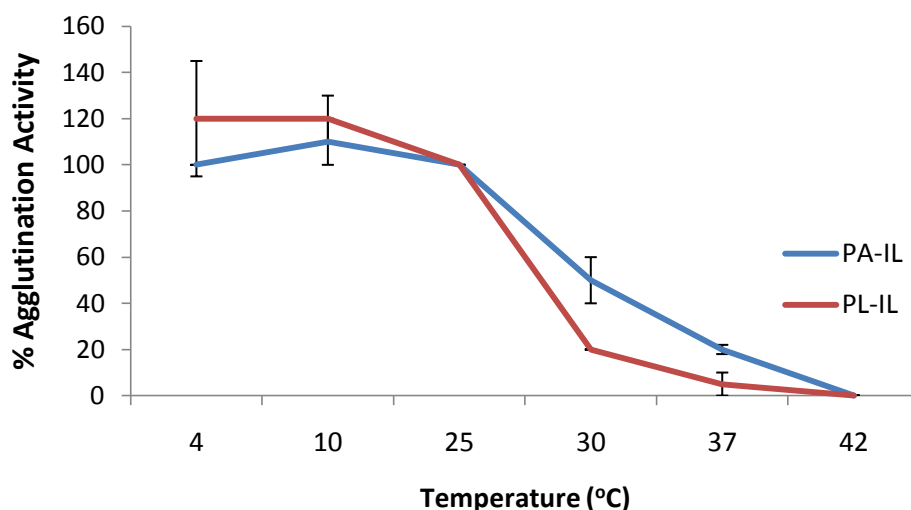


Fig 5.2 Effect of temperature on hemagglutination activity. Hemagglutinations were carried out at various different temperatures using different quantities of lectin. One HU is the quantity of lectin required for agglutination at room temperature. When less or more than this quantity is required for agglutination, the percentage activity was calculated.

The result that PA-IL and PL-IL readily agglutinate rat RBCs, while PA-IL_{mut1} does not, immediately indicates a drastic difference in specificity/affinity between the molecules. Secondly, it is also clear that PA-IL and PL-IL share clear similarities in their preference for galactose containing molecules (Table 5.1), and in their preference for oligosaccharides over monosaccharides (Fig 5.1), as low concentrations of these sugars are required to inhibit hemagglutination. PA-IL, in agreement with previous studies (Chen *et al.*, 1998) has the highest preference for hydrophobic derivatives of galactose, namely phenyl- β Gal and nitrophenyl- β Gal. Raffinose in contrast preferred raffinose which has the structure Gal α (1-6)Glc β (1-2)Fru, stachyose, which is composed of Gal α (1-6)Gal α (1-6)Glc β (1-2)Fru and melibiose which is Gal α (1-6)Glc. A common feature of these oligosaccharides is that they terminate in alpha-linked galactose. Both raffinose and stachyose share the terminal Glc β (1-2)Fru structure, commonly known as sucrose.

Both lectins show a very low affinity for sucrose, so it can be concluded that it is the α (Gal) linkage to which they are displaying an affinity.

It was also shown that the lectins PA-IL and PL-IL agglutinate red blood cells more efficiently at low temperatures (Fig 5.2), with both lectins displaying a significant drop in hemagglutination activity above 25°C. It is unknown whether lectin affinity is increased at this temperature due to a change in its structure, or if the rat erythrocytes undergo a physical change whereby glycan units are more available to the lectin.

5.3 Establishment of the enzyme linked lectin assay (ELLA)

The principle of the ELLA has been outlined in Section 1.8.1, but in summary the ELLA is a modified ELISA where the primary antibody has been replaced by a lectin, and the protein target is a glycoprotein (McCoy *et al.*, 1983; Kirkeby *et al.*, 2002). In this study, a (His)₆ affinity tag that serves as a purification tool and a target for a secondary antibody has been engineered into the lectin. The effects of this tag have already been established at a structural level (Chapter 4), and from preliminary hemagglutination analysis it is known that though this tag did affect valency, it did disrupt sugar-binding activity (Table 4.7). If the (His)₆ tag was shown to destroy or reduce sugar binding activity, other tags such as strepII biotin could similarly have been engineered into the lectin. Lectins are most commonly tagged with biotin, with many biotinylated lectins now commercially available. One study also described the labelling of lectins with a radioactive isotope such as Na¹²⁵I and the binding event detected by measuring radioactivity (Spiro *et al.*, 1984).

All lectin, glycoprotein and secondary antibody concentrations used in the ELLA are outlined in Section 2.24, and any deviations from this method are highlighted.

A number of different blocking agents and detergents were investigated to maximise the positive signal emitted by the lectin in the assay. From the literature it has already been proven that optimisation was necessary for the ELLA (Kim *et al.*, 2008) as some lectins used were found to bind to the blocking agent BSA. As the lectin is usually diluted in a 1% (w/v) solution of the blocking agent prior to

incubation in the assay (See Section 2.24), any interaction that is occurring between the lectin and BSA would severely quench any binding signal. As a result, a variety of blocking agents and detergents were investigated in terms of the binding of PA-IL and PL-IL.

In lectin arrays (eg Procognia Ltd lectin array) or other lectin assay systems, glycoproteins are often denatured. Therefore the effect of glycoprotein denaturation in ELLAs was investigated (Fig 5.4). In this assay, a range of glycoproteins were denatured by heating, and the binding of three lectins (PA-IL, PL-IL and ConA) to native and denatured glycoproteins was compared.

Many commercially available plant lectins such as ConA are employed in this study. These lectins are bioinylated, and their binding in the ELLA detected through the use of a HRP-labelled anti-biotin antibody. The specificities of all plant lectins are provided in Appendix A.

Calcium has already been shown to be necessary for the function of PA-IL (Cioci *et al*, 2003). This is due to the location of the ion in the sugar binding pocket, where it makes direct contact with the galactose residue upon binding. The chemical compound ethylenediaminetetraacetic acid (EDTA) is widely used to sequester di- and trivalent metal ions. When used in the ELLA format, the dependence of the lectins on metal ions can easily be elucidated. To investigate the role of calcium in PA-IL, PL-IL and PA-IL_{mut1} binding lectins were incubated with 20 mM EDTA for two hours prior to buffer exchanging into TBS, and subsequent incorporation into the ELLA assay. As shown in Fig 5.5, incubation of the three lectins with EDTA significantly inhibited the binding of each lectin indicating the necessity of calcium for binding

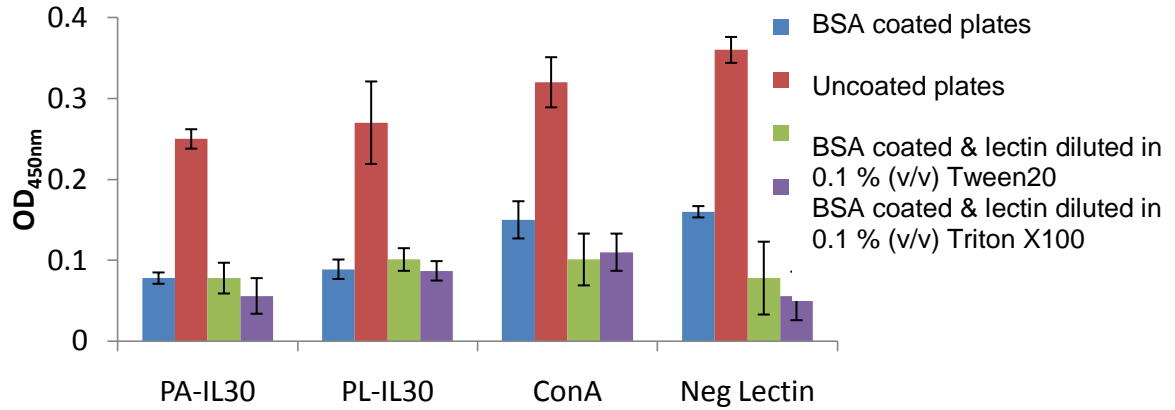


Fig 5.3 Comparison of blocking reagents in ELLA. Quantitative detection of lectin bound to BSA coated and uncoated plates assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). The binding of PA-IL30, PL-IL30 and the control lectin ConA to differently blocked empty wells in the Nunc ELISA plate is compared.

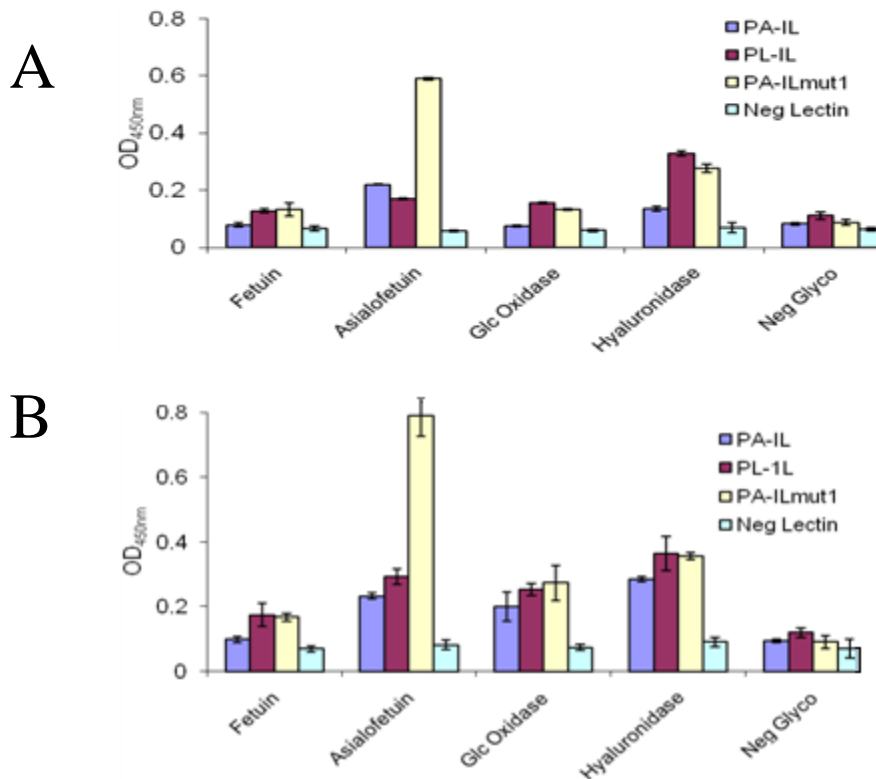


Fig 5.4 Comparison of native and denatured glycoprotein in ELLA. Quantitative detection of lectin bound to various glycoproteins assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). The effect of native immobilisation (A) and denaturing immobilisation (B) on lectin binding signal was compared. Three recombinant lectins were used to evaluate lectin binding. Slightly higher signals were seen when lectins were incubated with denatured glycoproteins, though the error also increased.

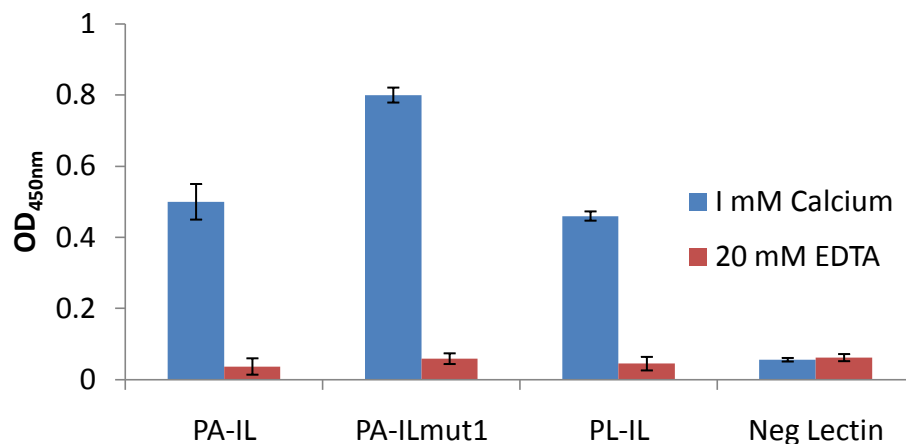


Fig 5.5 Calcium is required for PL-IL and PA-IL_{mut1} binding. Quantitative detection of lectins bound to thyroglobulin assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). The effect of the removal of calcium ion from the lectin binding pocket by EDTA on lectin binding to the glycoprotein porcine thyroglobulin was evaluated.

It was discovered the coating the ELLA well with BSA significantly lowered the background binding signal within the ELLA assay, which could be further reduced by the addition of detergents. It can be seen from Fig 5.3 that the detergent Triton X100 proved slightly more effective than Tween 20, while blocking of the well with BSA was preferable to not having a blocking step, as the binding signals were consistently lowest for PA-IL, PL-IL and a positive control lectin, ConA when BSA coated plates were used in conjunction with Triton X 0.1 (% v/v). The effect of glycoprotein denaturation was also investigated. It was concluded that denaturation of the glycoprotein did slightly increase the binding signal, however the error also increased (Fig 5.4) and consequently, in future assays glycoprotein denaturation was not employed. Finally, it was confirmed that like PA-IL, PL-IL and PA-IL_{mut1} have a dependency on a metal ion which is removed through the addition of EDTA (Fig 5.5).

5.4 Quaternary structure affects lectin detection in ELLA

In Chapter 4 it was determined that the position of the (His)₆ tag could disrupt quaternary structure of PA-IL, but not PL-IL. It is important to understand the degree of multi-valency of a lectin when investigating the affinity of a lectin, as this is directly related to its valency. This relationship is even more important in the ELLA method, as the detection of the lectin-binding event is dependent on the binding of a secondary antibody to a ligand present on the lectin molecule. If the lectin is tetrameric, there will be four secondary antibody targets available which will indicate the binding of one molecule, compared to only two for a dimer. Consequently, affinities of tetramers and dimers cannot be directly compared using the method. As a result it would have been desirable to use the PA-IL60 and PL-IL60 molecules in the assay, as both formed tetramers according to Chapter 4. However, though the (His)₆ tag did not disrupt the functionality of the lectin according to the hemagglutination assay (Section 4.10), the binding of the (His)₆ tag was not detectable by interrogation with an HRP-linked α -(His)₆ antibody when located on the C-terminus of the lectin (Fig 5.6). However, positioning the (His)₆ tag on the N-terminus allowed the secondary antibody to bind the (His)₆ tag and therefore the presence of the lectin could be detected.

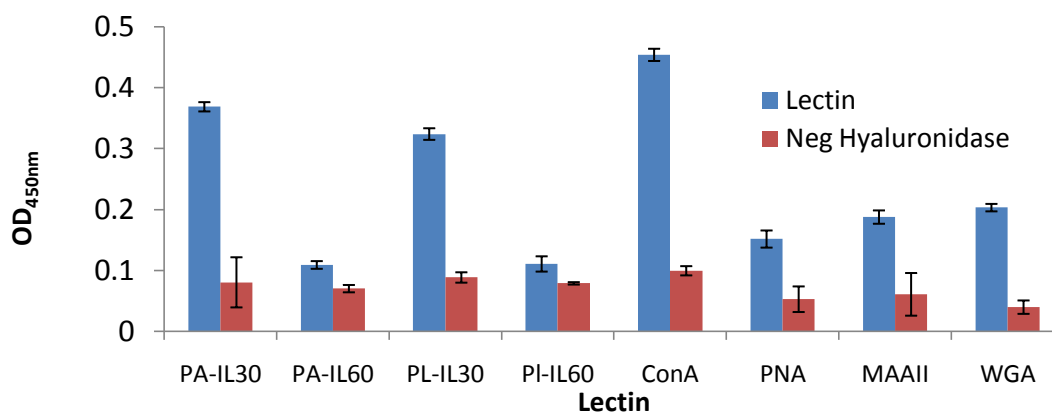


Fig 5.6 Binding of differently tagged forms of PA-IL and PL-IL to immobilised glycoprotein (hyaluronidase). Quantitative detection of lectins bound to immobilised hyaluronidase assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). Analysis of the binding of N-terminally and C-terminally (His)₆ tagged PA-IL and PL-IL to immobilised hyaluronidase. The commercial lectins ConA, PNA, MAAII, and WGA which bind to mannose, galactose, sialic acid and GlcNAc respectively are employed as positive controls in the assay.

It was discovered, that though *N*-terminally tagged and *C*-terminally PA-IL and PL-IL were both active in the hemagglutination assay (Table 4.7), the *C*-terminally tagged versions proved to show no activity in the ELLA (Fig 5.6). The (His)₆ tag was exploited in Section 3.4 for the purification of these *C*-tagged molecules by IMAC, so it is known that the tag was present. It was concluded that although the presentation of the tag in both PA-IL60 and PL-IL60 permitted IMAC purification, it was not sufficiently exposed for the binding of a α -(His)₆ antibody. Consequently, in all subsequent ELLA assays the PA-IL30 and PL-IL 30 molecules were used for the determination of the specificity of PA-IL and PL-IL. This unfortunately means that affinity results for the tetrameric PL-IL and the PA-IL could not be compared. However, it was shown that PA-IL30 and PA-IL_{mut1}30 display similar quaternary structures (Section 4.7), and these molecules can be directly compared.

5.5 Investigation into recombinant lectin specificity by ELLA

Having optimised the conditions for ELLA in Section 5.3, the specificity of the lectins PA-IL, PL-IL and PA-IL_{mut1} were investigated by ELLA. In section 5.2 it had already been determined that the lectins PA-IL and PL-IL have a high affinity for complex galactose containing oligosaccharides and derivatised galactose molecules. Galactose is a commonly occurring structure in *N*-glycan structures which are displayed on many glycosylated proteins, though it is usually capped by the terminal sugar, sialic acid. The binding of the recombinant lectins to an array of commercially available glycoproteins was investigated with only thyroglobulin showing any significant binding for the three lectins (Fig 5.7 A, 5.7B and 5.7C). Next, the glycoproteins were treated with the enzyme neuraminidase, which removes terminal sialic acid, and should expose the underlying galactose residues. The glycosidase reaction was confirmed through the binding of sialic acid specific

commercial lectins in a separate ELLA (Fig 5.9). The removal of sialic acid resulted in increased reactivity of recombinant PA-IL, PL-IL and PA-IL_{mut1} to the glycoproteins fetuin, thyroglobulin, hyaluronidase and ovalbumin (Fig 5.8). Having created some target glycoproteins for which the recombinant lectins display an affinity, the specific sugars in the underlying glycan structures were then sequentially removed from these glycoproteins by glycosidases, to discover exactly which linkages the lectins were binding (Fig 5.10 and Fig 5.11). Commercial biotinylated lectins were used as positive controls for the glycosidase reactions to ensure the sugar removal was complete and specific, as these enzymes can be unstable and display broad specificity. The specificity of these commercial lectins are given in Appendix B. Fig 5.10 shows the binding signal for PA-IL against fetuin increase upon neuraminidase treatment, a fall after α -galactosidase treatment, and no change after β -galactosidase treatment. PL-IL shows an increase in signal upon neuraminidase treatment, a fall after α -galactosidase treatment, and a smaller fall after β -galactosidase treatment. PA-IL_{mut1} displays a starkly different binding profile to that of PA-IL as there is an increase upon neuraminidase treatment, no change after α -galactosidase treatment and a decrease after β -galactosidase treatment. Fig 5.12, which displays the same set of glycosidase treatments to the glycoprotein porcine thyroglobulin, shows a broadly similar trend, with PA-IL and PA-IL_{mut1} displaying contrasting specificities.

The sources of the glycoproteins are as follows; fetuin, bovine; thyroglobulin, porcine; hyaluronidase, bovine; ovalbumin, chicken; glucose oxidase, *Aspergillus niger*; invertase, bakers yeast; transferrin, human; RNase B, bovine; carbonic anhydrase, bovine; and urease from the jack bean.

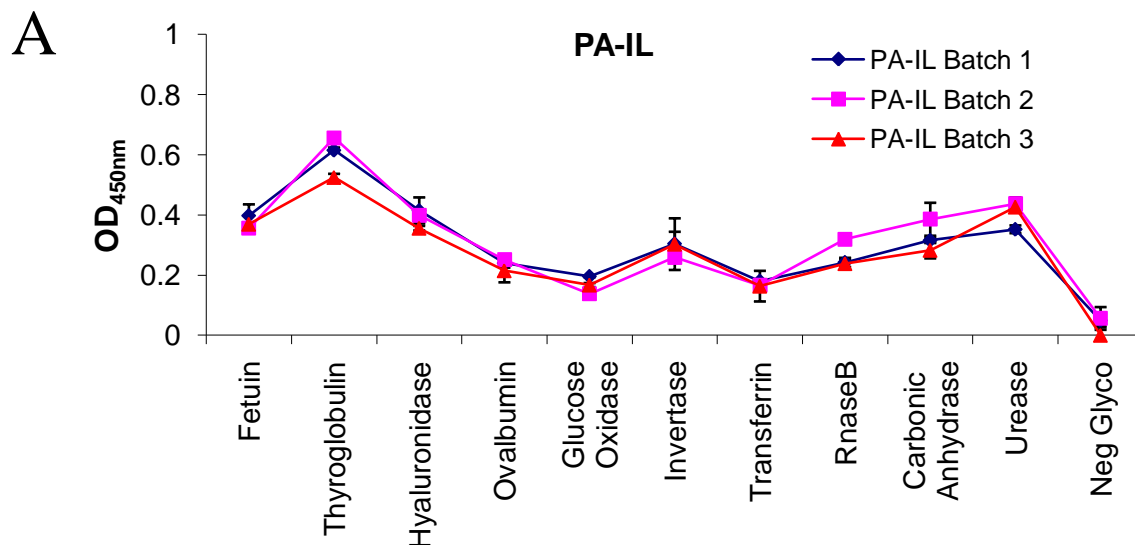


Fig 5.7A Identification of target glycoprotein molecules for PA-IL by ELLA. Quantitative detection of PA-IL bound to a range of commercially available glycoproteins assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). The three lines represent three purified batches of PA-IL from three separate fermentations.

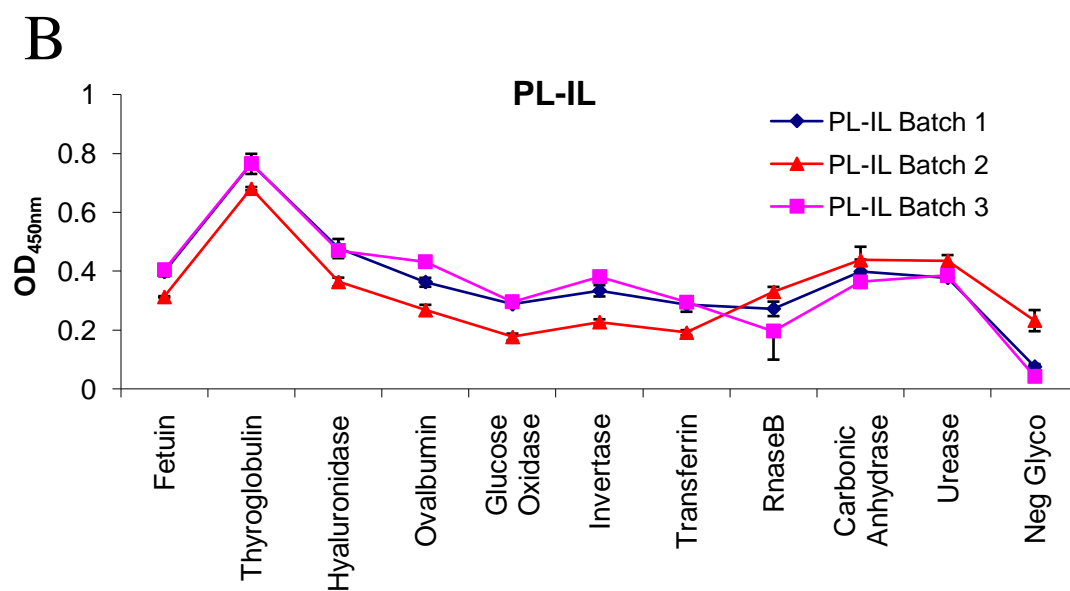


Fig 5.7B Identification of target glycoprotein molecules for PL-IL by ELLA. Quantitative detection of PL-IL bound to a range of commercially available glycoproteins assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). The three lines represent three purified batches of PL-IL from three separate fermentations.

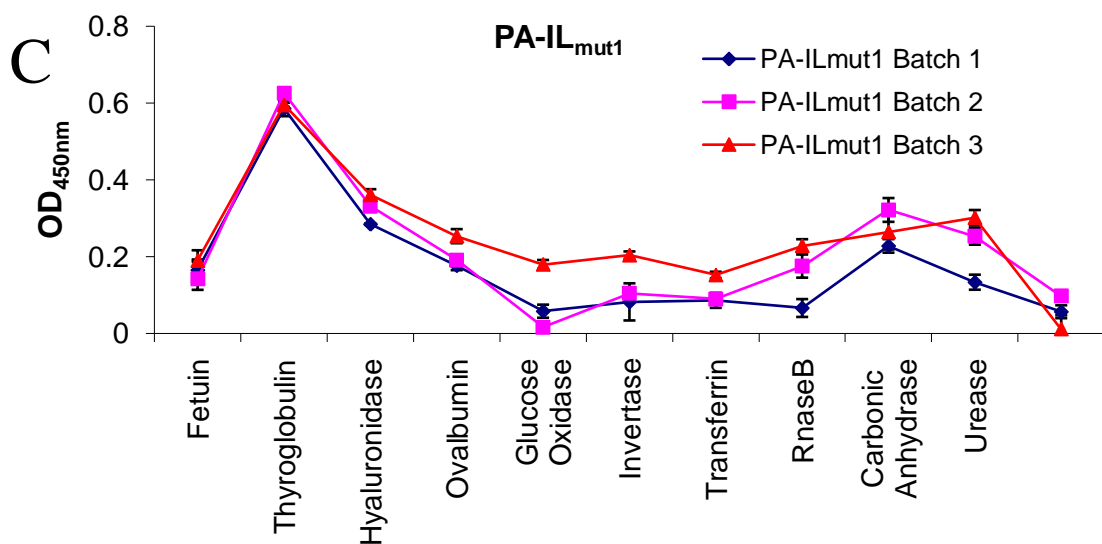


Fig 5.7C Identification of target glycoprotein molecules for PA-IL_{mut1} by ELLA. Quantitative detection of PA-IL_{mut1} bound to a range of commercially available glycoproteins assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). The three lines represent three purified batches of PA-IL_{mut1} from three separate fermentations.

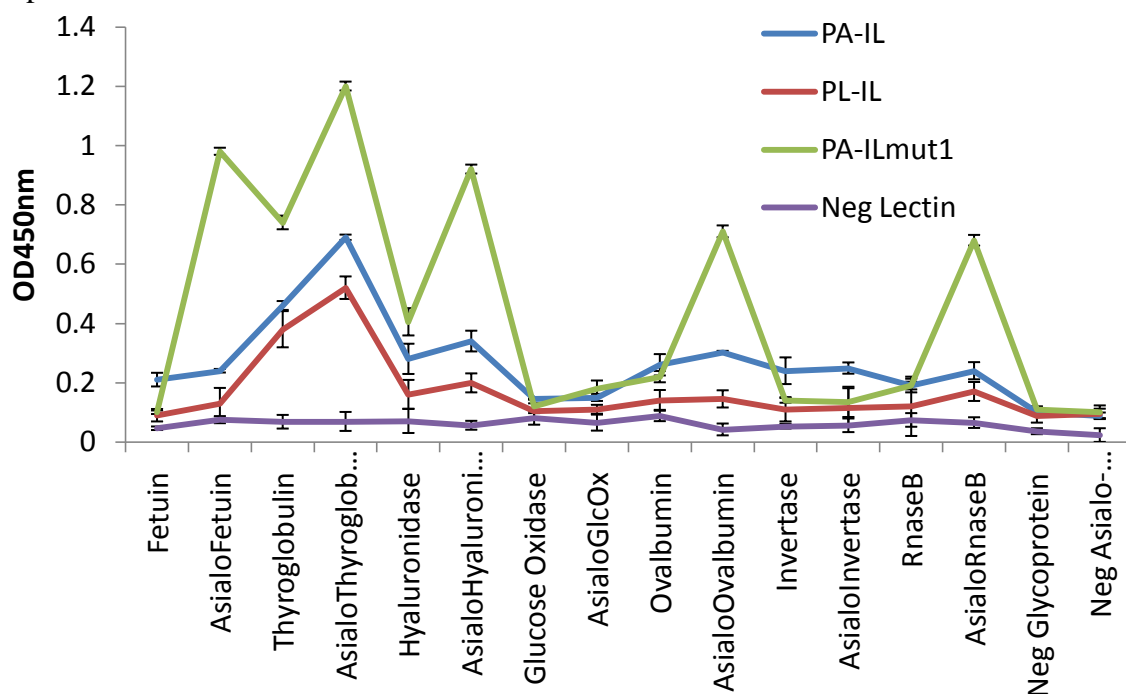


Fig 5.8 Investigation of recombinant lectin binding to asialylated glycoproteins. Quantitative detection of recombinant lectins bound to a range of glycoproteins treated with neuraminidase and assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). Lectin binding to un-treated versions of the same glycoproteins was also investigated. Neuraminidase (Sigma-Aldrich) was used as per manufacturer's instructions.

Fig 5.8 shows that the relatively low signals seen in Fig 5.7A, 5.7B and 5.7C which represented PA-IL, PL-IL and PA-IL_{mut1} binding to a range of commercially available glycoproteins could be significantly increased upon removing the sugar sialic acid from the glycoproteins. In Section 1.2, it was outlined how sialic acid is often the terminal sugar on a glycan chain, and its removal by the enzyme neuraminidase will expose many underlying glycans, the most common of which is galactose. The binding of PA-IL and PL-IL to asialylated version of glycoproteins (neuraminidase treated glycoproteins) was significantly increased for thyroglobulin and hyaluronidase, while PA-IL_{mut1} binding increased with respect to all glycoproteins except for glucose oxidase and invertase. These glycoproteins were purified from *Aspergillus niger* and bakers yeast respectively organisms that would not be expected to decorate glycans with galactose residues (Section 1.2 and Section 6.2.5).

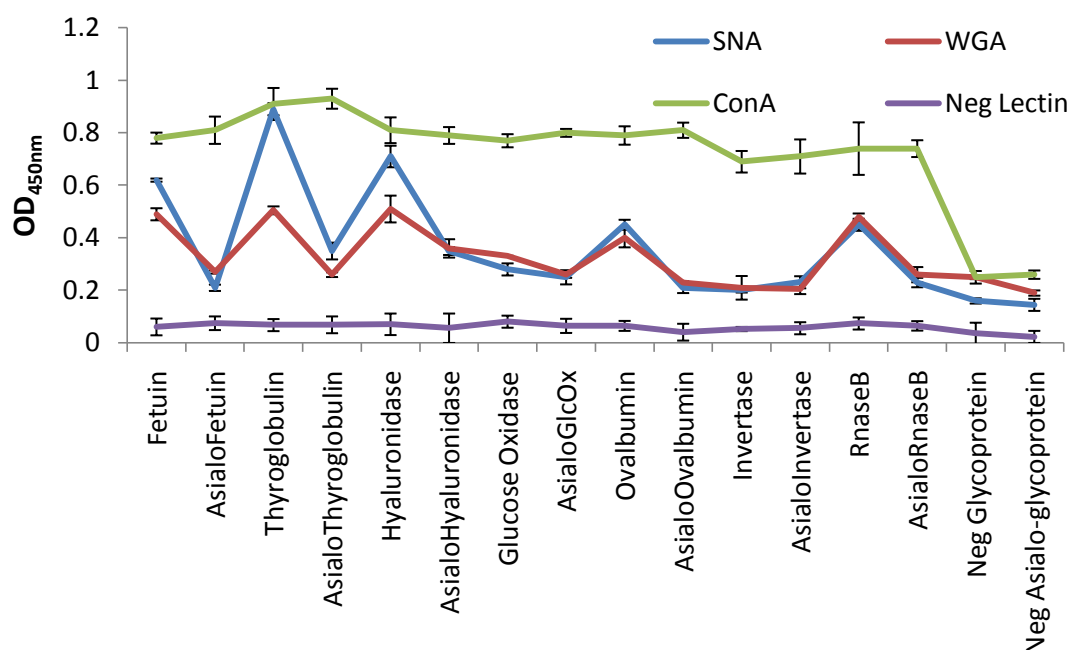


Fig 5.9 Binding of recombinant lectins to neuraminidase treated glycoproteins. Quantitative detection of biotinylated lectins bound to a range of neuraminidase treated glycoproteins assayed by the activity of an enzyme-linked secondary antibody (OD_{450nm}). Neuraminidase (Sigma-Aldrich) was used as per manufacturers instructions, and immobilised. The ability of ConA, WGA and SNA to bind to these glycoproteins was then used to elucidate if neuraminidase treatment was successful.

The lectin SNA binds specifically to $\alpha(2-6)$ linked sialic acid. WGA binds to sialic acid, but also displays an affinity for the residue GlcNAc, which is a common glycan in *N*-linked structures (Section 1.2). If the neuraminidase treatment of the various glycans has been successful, the binding signal for SNA should be high for a glycoprotein, and disappear once it has been neuraminidase treated. Similarly, the signal should be high for WGA, and should disappear upon sialic acid removal unless the sugar GlcNAc has been exposed. In Fig 5.9, it is clear that sialic acid removal occurred to completion, as the signal for SNA returned to baseline for each asialoglycoprotein except thyroglobulin and hyaluronidase, and in both cases the majority of signal has been removed (baseline is determined by SNA binding to a negative glycoprotein sample in the last two columns). ConA is another control lectin that binds to the core mannose structure present in all *N*-linked glycan chains. It serves as a control to ensure the same amount of glycoprotein has been immobilised in each instance. In this assay, ConA signals remain constant between each glycoprotein and corresponding asialoglycoprotein, indicating that any reduction in SNA/WGA signal was not due to a reduction in the amount of glycoprotein that has been immobilised. This assay serves a useful purpose as when interpreting the data shown in Fig 5.8, as it can be concluded that the increase in signals of PA-IL, PL-IL and PA-IL_{mut1} for many of the glycoproteins were as a result of the removal of sialic acid which exposed some glycan residue to which each lectin bound. The identification of these glycans was the subject of further investigation.

Next, it was attempted to further define the glycans within the glycan chains to which PA-IL, PL-IL and PA-IL_{mut1} display an affinity. In *N*-linked structures, sialic acid usually is attached to a galactose molecule, so two galactose enzymes were used to differentiate between α -linked and β -linked galactose, both of which have been found in *N*-linked glycan chains.

Asialo-glycoproteins were treated with α -galactosidase (source green coffee beans) and $\beta(1-4)$ -galactosidase (recombinant enzyme purified from *E. coli*) and immobilised as per previous ELLA experiments. A number of biotinylated lectins

were used in conjunction with the recombinant lectins to act as controls for the glycosidase treatments, and to identify the sugars present on the glycoprotein. The two glycoproteins investigated in this manner were fetuin and thyroglobulin.

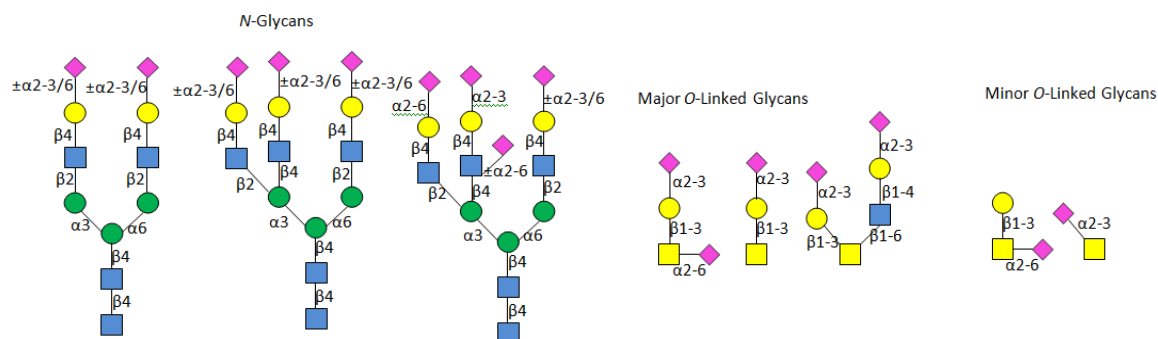


Fig 5.10 Glycan structures that have been found on fetuin. The structures and linkages of the major *N*-linked and *O*-linked glycans that were found to be present on fetuin (Iskratsch *et al.*, 2009). Nomenclature for the glycans can be found in Appendix A.

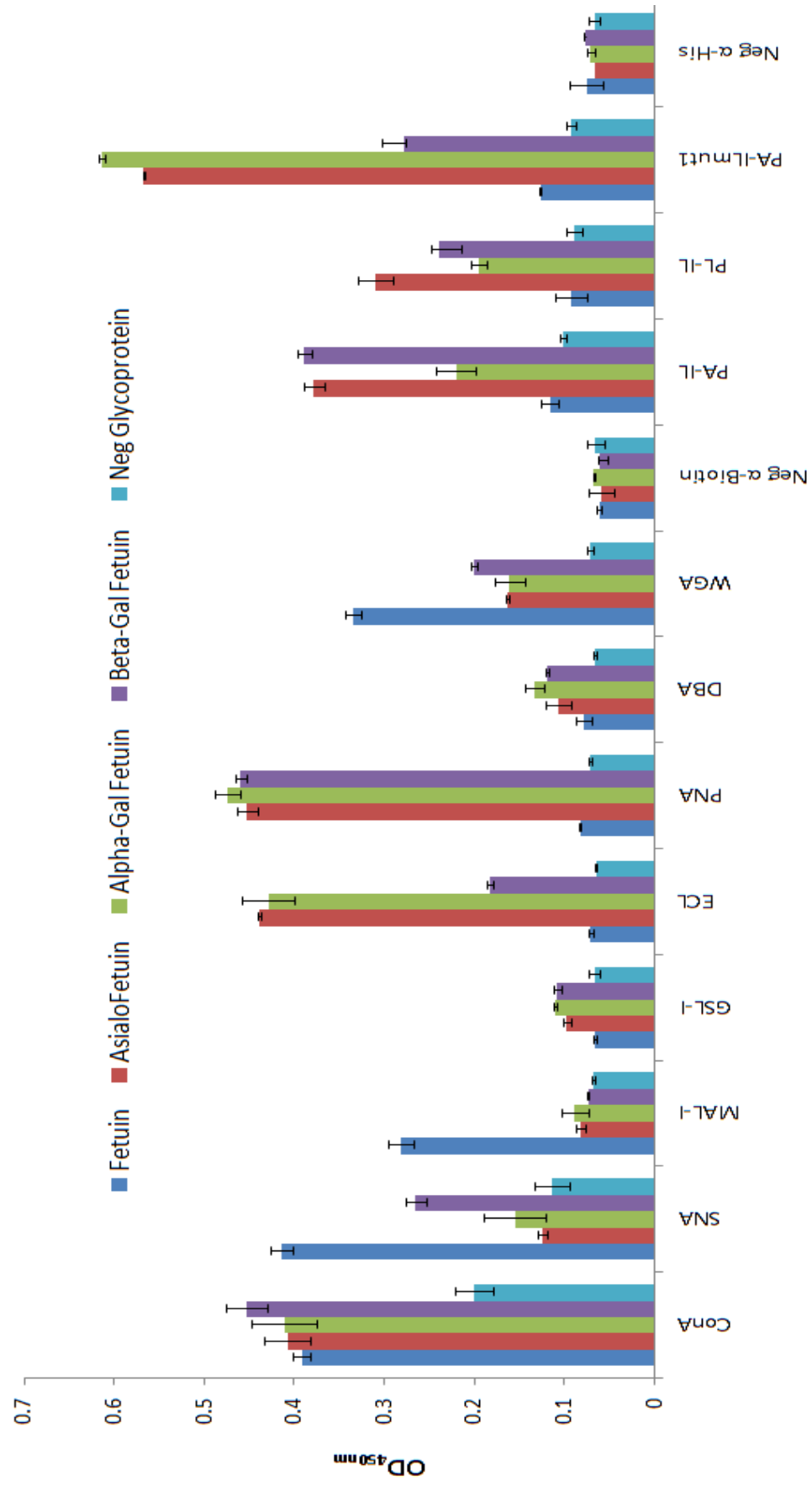


Fig 5.11: Determination of recombinant lectin specificity by analysis of binding to glycosidase treated fetuin. Quantitative detection of biotinylated and recombinant (His)₆ tagged lectins bound to a glycosidase derivatives of fetuin assayed by the activity of an enzyme-linked secondary antibody (OD_{450nm}). These glycosidases include neuraminidase, α-galactosidase, and β(1-4) galactosidase. The specificity of the commercial lectins is listed in Appendix A.

When the *N*-linked glycans of fetuin are examined, the only structure that is exposed upon neuraminidase treatment of fetuin is $\beta(1-4)$ linked galactose (Fig 5.10). However the *O*-linked structures display higher variability, with several other possible residues exposed to which each lectin could display an affinity.

When the binding profiles of biotinylated plant lectins are investigated in Fig 5.11, a great deal of information about the glycans present on the glycoprotein can be extracted. ConA, which acts as an internal positive control, displays a positive binding signal at the same level for each derivative of fetuin. This control shows the same amount of glycoprotein was immobilised in each case. It is important to highlight at this stage that ConA is the only lectin used that binds to core glycan structures. All other biotinylated lectins bind to terminal sugars only. SNA binds to sialic acid. After neuraminidase treatment of fetuin, the binding signal for this lectin drops to baseline. The WGA lectin is known to bind sialic acid and GlcNAc, and its signal can be seen to drop upon neuraminidase treatment and rise again upon β -galactosidase treatment, as GlcNAc is usually exposed upon the removal of $\beta(1-4)$ galactose.

The lectins GSL-I and DBA show no binding to any derivative of fetuin, indicating that the sugars α -linked galactose and GalNAc are not present in significant quantities on the glycoprotein, which disagrees slightly with the structures identified by Iskratsch *et al.*, 2009 (Fig 5.10), as these sugars are proposed to exist on the *O*-linked glycans. It is possible however that signal quenching by BSA is occurring for these lectins.

The PNA and ECL binding patterns to fetuin derivatives prove very informative in determining the specificity of the three recombinant lectins. PNA is specific to terminal β -linked galactose, and as expected, the lectin doesn't bind to fetuin, a large signal is seen for asialofetuin, and the signal does not drop after α -galactosidase or $\beta(1-4)$ galactosidase treatment, indicating the presence of a $\beta(1-3)$ galactose on the molecule to which the lectin is binding. The ECL lectin is specific for $\beta(1-4)$ linked galactose, and the signal for this lectin drops significantly after $\beta(1-4)$ galactosidase treatment, though not completely, indicating incomplete activity of the enzyme.

The PA-IL_{mut1} lectin shows an identical binding pattern to the ECL lectin, suggesting that this too is a $\beta(1-4)$ galactose specific lectin. This is in contrast to PA-IL, which it is known binds to $\alpha(1-3)$ and $\alpha(1-4)$ linked galactose (Section 1.9.3). When the fetuin binding pattern of PA-IL is investigated it is clear that the lectin displays an affinity for some glycan exposed after neuraminidase treatment. This binding signal is then reduced upon α -galactosidase treatment. This is an unexpected result as there are no reported of any α -galactose structure on fetuin, suggesting an α -galactose containing contaminant present within fetuin, or that α -galactosidase has non-specifically cleaved the glycan to which PA-IL is binding, e.g GalNAc. It has been shown previously the PA-IL displays a low specificity for β -linked galactose structures (Section 5.2), however if PA-IL was binding to these glycans in fetuin, the signal would have reduced upon β -galactosidase treatment and not upon α -galactosidase treatment.

PL-IL has an unknown specificity, but like PA-IL, it is clear that the lectin's ability to bind fetuin is increased upon neuraminidase treatment of the glycoprotein suggesting an exposed galactose residue is being bound. The binding of the molecule to asialofetuin is then decreased upon both α -galactosidase and β -galactosidase treatment of the glycoprotein, indicating a broader specificity of the lectin.

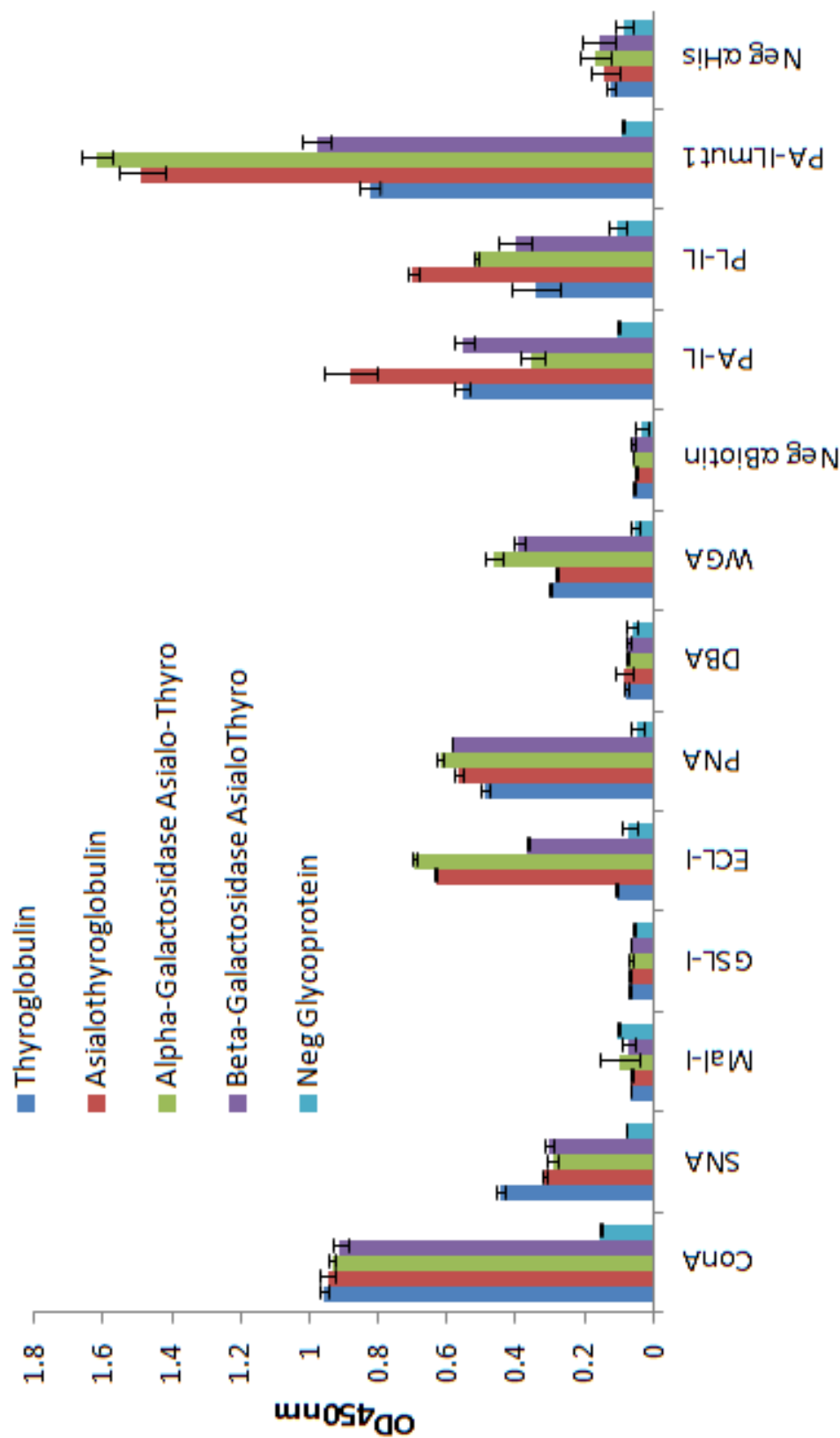


Fig 5.12 Determination of recombinant lectin specificity by analysis of binding to glycosidase treated thyroglobulin
 Quantitative detection of biotinylated and recombinant (His)₆ tagged lectins bound to glycosidase-treated derivatives of thyroglobulin assayed by the activity of an enzyme-linked secondary antibody (OD_{450nm}). These glycosidases include neuraminidase, alpha-galactosidase, and beta(1-4) galactosidase. The affinity of the commercial lectins is listed in Appendix A.

Unlike fetuin, there has not been a comprehensive glycoprofiling study for porcine thyroglobulin, partly due to the protein having 15-20 different *N*-glycosylation sites with its glycan chains comprising 10% of its dry weight. However studies such as Rosenfeld *et al.*, (2007) and Kamerling *et al.*, (1988) have identified many of the individual sugar residues that exist on the glycoprotein. The presence of α -linked galactose and GalNAc was described in these studies, and it was proposed that these occurred in *O*-linked glycans.

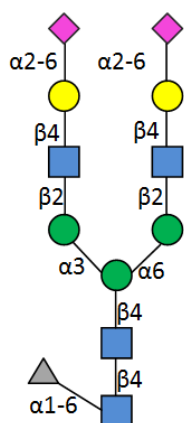


Fig 5.13 The structure of the most abundant *N*-linked glycan chain present in porcine thyroglobulin. Cartoon representation of the glycan assembly in the most common *N*-linked structure found on porcine thyroglobulin (Kamerling *et al.*, 1988). Nomenclature is as described in Appendix A.

Like the fetuin profiling experiment, the binding profile of ConA is examined first, to ensure that equal amounts of each thyroglobulin derivative have been immobilized. The SNA signal, has not returned to baseline after neuraminidase treatment, indicating incomplete cleavage of sialic acid, or the presence of sialic acid linked to another sugar by a linkage to which it could not cleave. However enough sialic acid was removed for the binding of the galactophilic lectins PNA and ECL to increase their respective binding signals to the asialo-thyroglobulin. Similarly to the fetuin experiment, α -galactosidase and β (1-4) galactosidase treatment did not result in a lower PNA binding signal. Again, β (1-4) galactosidase did not completely reduce the signal of the β (1-4) galactose specific ECL lectin.

Again, PA-IL_{mut1} displayed a similar binding profile to ECL, which suggests that the specificity of this molecule is β (1-4)galactose. This is an important result, as it shows that by altering just three amino acids within the sugar binding pocket of PA-IL, the specificity of the molecule was completely altered.

PA-IL, as expected binds to asialo-thyroglobulin strongly with a significant reduction in signal after α -galactosidase treatment confirming that the molecule is strongly α -galactophilic. PL-IL once again shows a reduced signal for both α -galactosidase and $\beta(1-4)$ galactosidase treatment, indicating a broad specificity for galactose containing molecules.

5.6 Investigation of recombinant lectin affinity

The affinity of one molecule for another is most commonly represented by a value termed the K_a (the association constant), or K_d (the dissociation constant $\frac{1}{K_a}$). The dissociation constant is more commonly used, because although two molecules can have the same affinity for a target, their rates of dissociation can be very different. Hence, the K_d is a more useful measurement, and is calculated from the amount of substrate that leads to half-maximal specific binding. The lower the K_d the higher the affinity as it will take a lower amount of lectin to bind half of the immobilised target proteins.

Equation 5.1

$$Y = \frac{P_{AMAX} \cdot X}{K_d + X}$$

In this equation Y represents the specific binding, P_{AMAX} is the absorption values at which the maximum number of binding sites is occupied, with X representing the concentration of the glycoprotein. It is difficult to determine the affinity of a lectin for an α -linked galactose glycan, as target molecules displaying this sugar are not widely commercially available. Of the commercial glycoproteins available in this study, only thyroglobulin has had the presence of α -galactose

previously reported, and this glycoprotein has been the subject of very limited glycoprotein studies. Fetuin, which lacks α -galactose, has been the subject of intense carbohydrate analysis, with its glycan profile having been extensively analysed, and is provided in Fig 5.10. For this reason it was useful for the study of PA-IL_{mut1}. The affinities for each lectin towards thyroglobulin and fetuin were calculated in Fig 5.14 and 5.15

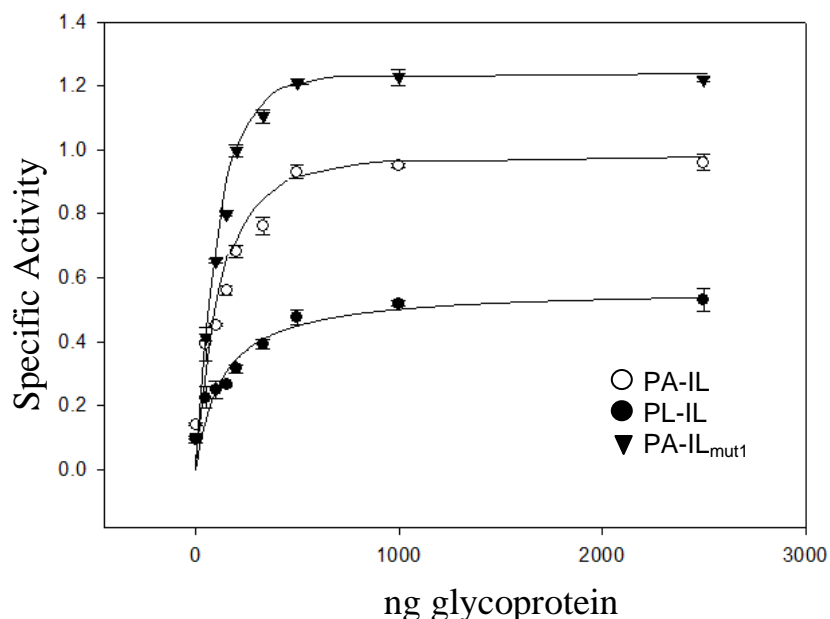


Fig 5.14 Determination of recombinant lectin affinity for asialothyroglobulin. Quantitative detection of recombinant lectin bound to various concentrations of asialothyroglobulin assayed by the activity of an enzyme-linked secondary antibody (OD_{450nm}). The constant lectin concentration is 7nM, with immobilized asialothyroglobulin concentrations ranging from 20 to 2500 ng concentration. P_{AMAX} values for PA-IL, PL-IL and PA-IL_{mut1} were 1.04, 0.57, and 1.35, and K_d values were 108 ng 129 ng and 94 ng respectively.

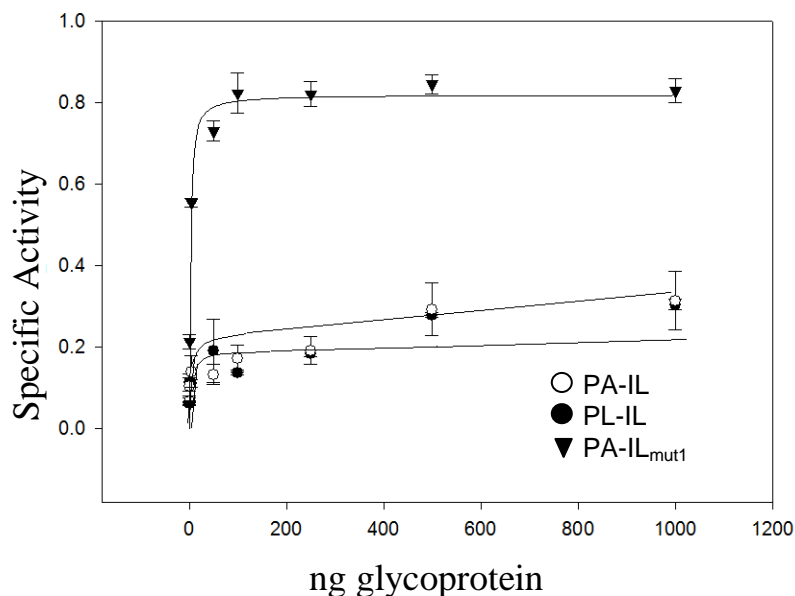


Fig 5.15 Determination of recombinant lectin affinity for asialofetuin. Quantitative detection of recombinant lectin bound to various concentrations of asialofetuin assayed by the activity of an enzyme-linked secondary antibody (OD_{450nm}). The constant lectin concentration is 7nM, with immobilized asialofetuin concentrations ranging from 20 to 1,000 ng concentration. P_{AMAX} values for PA-IL, PL-IL and PA-IL_{mut1} were 0.31, 0.22, and 0.82, and K_d values for PA-IL and PL-IL were not detectable within this range, but PA-IL_{mut1} was 2.43 ng.

Depending on the glycoproteins used, different K_d values for each lectin were obtained. In each case PA-IL_{mut1} had the lowest K_d indicating the highest affinity. From the characterization results displayed in Fig 5.11 and 5.12, it is presented that the sugar to which PA-IL_{mut1} has a high specificity for is $\beta(1-4)$ galactose, an abundant sugar on *N*-linked structures, so this is the expected result. PL-IL, in Fig 5.11 and Fig 5.12 was shown to have the broadest specificity of the recombinant lectins, as it could not be ascertained exactly to which glycan structure it was binding. This low specificity to the displayed glycans is possibly contributing to a high K_d value, whereas the high specificity of PA-IL_{mut1} to $\beta(1-4)$ galactose residues is a contributing factor for its low K_d values.

Competition ELLAs were carried out using asialothyroglobulin as the target glycoprotein, to further expand the understanding of each lectin's specificity. The competing sugars used are galactose, raffinose, melibiose, glucose, phenyl- β -galactose, nitrophenyl- β -galactose, cellobiose and lactose, whose structure can be seen in Appendix A. When interpreting this assay, it is important not to compare the different inhibition profiles of each lectin, as it has been shown that in each case they are binding to a different target molecule on asialothyroglobulin (Fig 5.12). Instead, the affinity of each lectin to the inhibitory molecule can be deduced. Finally, PL-IL in this assay is a tetramer, and while present in the same concentration as the dimeric PA-IL and PA-IL_{mut1}, will dissociate more slowly from the glycoprotein, as it will have more binding site per molecule that will need to be occupied by the saccharide.

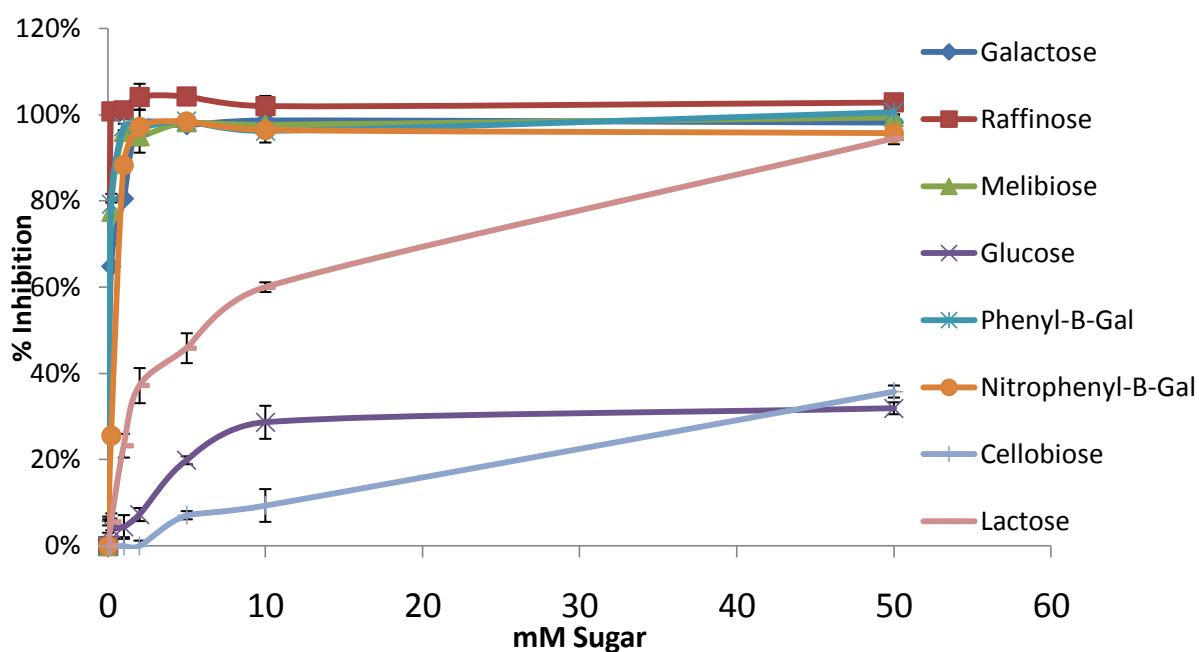


Fig 5.16 Inhibition of PA-IL binding to asialothyroglobulin by various sugars. Quantitative detection of PA-IL bound to a constant concentration of asialothyroglobulin in the presence of varying amounts of inhibitory sugars assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). The constant lectin concentration is 7nM, with inhibitory sugar concentrations ranging from 0 to 50 mM concentration. 100% binding was determined by incubation of asialothyroglobulin with PA-IL without competing sugar.

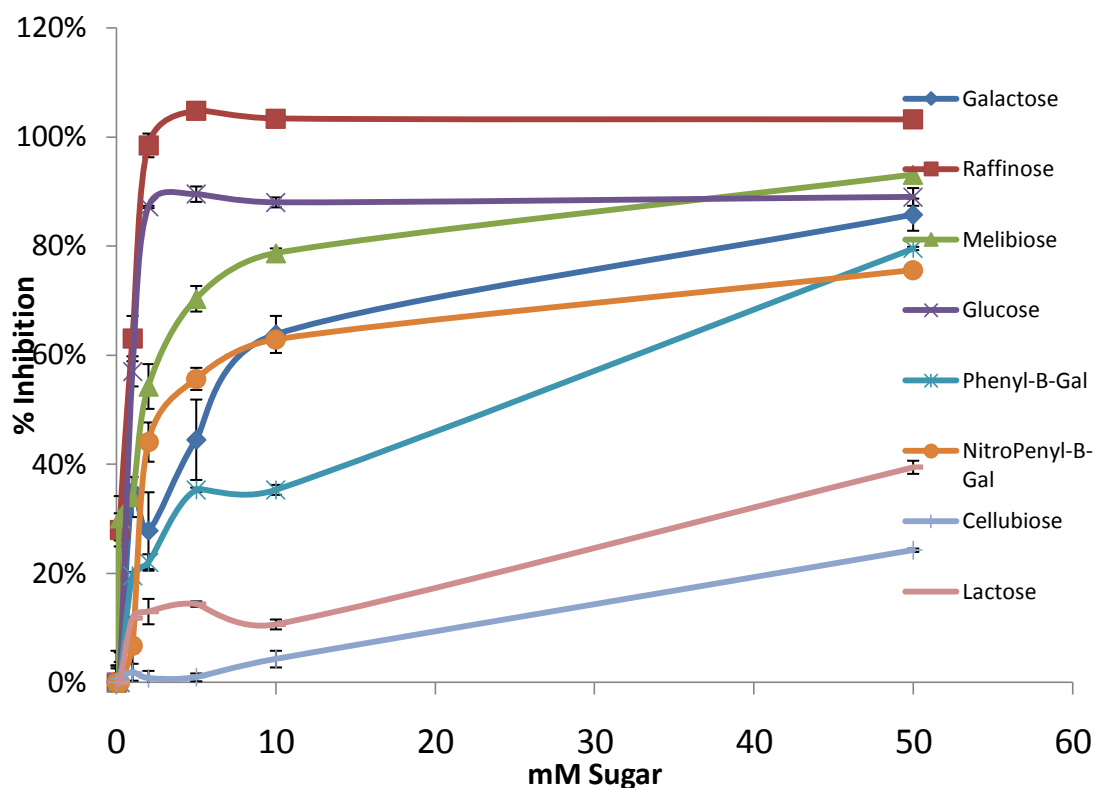


Fig 5.17 Inhibition of PL-IL binding to asialothyroglobulin by various sugars. Quantitative detection of PL-IL bound to a constant concentration of asialothyroglobulin in the presence of varying amounts of inhibitory sugars assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). The constant lectin concentration is 7nM , with inhibitory sugar concentrations ranging from 0 to 50 mM concentration. 100% binding was determined by incubation of asialothyroglobulin with PL-IL and no competing sugar.

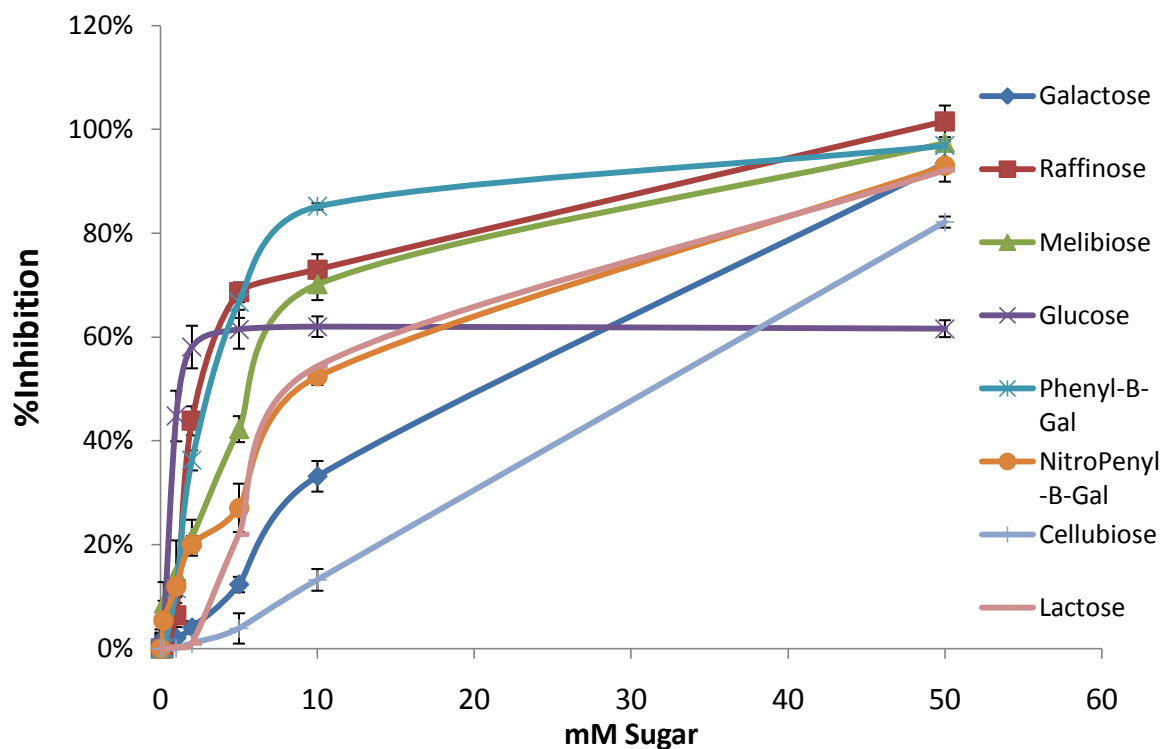


Fig 5.18 Inhibition of PA-IL_{mut1} binding to asialothyroglobulin by various sugars. Quantitative detection of PA-IL_{mut1} bound to a constant concentration of asialothyroglobulin in the presence of varying amounts of inhibitory sugars assayed by the activity of a HRP-linked secondary antibody (OD_{450nm}). The constant lectin concentration is 7nM, with inhibitory sugar concentrations ranging from 0 to 50 mM concentration. 100% binding was determined by incubation of asialothyroglobulin with PA-IL_{mut1} with no competing sugar.

In the competition assays presented in Figs 5.16, 5.17 and 5.18, very different inhibition profiles for each lectin can be seen. PA-IL inhibition reaches maximum at low concentrations of each sugar except for lactose, glucose and cellobiose. The predominant glycan type in thyroglobulin is *N*-linked structures, which generally lack the α -galactose residue. As a result it will not require large amounts of competing sugar to inhibit lectin binding to the immobilized glycoprotein. Secondly the lectin PA-IL30, exists in a dimer in its native state (Table 4.6), and consequently requires less sugar than the tetrameric PL-IL30 for inhibition. The most inhibitory oligosaccharide for PL-IL was raffinose followed by melibiose (in agreement with Section 5.2). The two least inhibitory sugars for PL-IL, like PA-IL were cellobiose and lactose. The monosaccharide glucose could inhibit the majority of binding at

low concentrations. From Fig 5.11, where it was suggested that PL-IL had a broad specificity, which correlates with the findings in this experiment, as glucose, galactose and melibiose could all bind some of the PL-IL molecule, but never completely inhibited the binding event.

The inhibitory profiles for PL-IL and PA-IL_{mut1} in terms of glucose inhibition proved very similar. PA-IL_{mut1}30, like PA-IL30, is a dimer, and as a result would require less sugar than PL-IL to inhibit binding. However, as seen from Fig 5.14, its affinity for asialothyroglobulin is much greater than the other two molecules. Consequently, far greater concentrations of sugar are required to inhibit the binding event. In comparison with the parent molecule PA-IL, glucose and lactose are more inhibitory, highlighting the change in specificity between the two molecules.

5.7 Application of recombinant lectins in the characterization of a biopharmaceutical product

Having proved the ELLA assay reproducible (Figs 5.7A, B and C) and useful in terms of obtaining useful data about the carbohydrate profile of a glycoprotein (Section 5.5), the use of the recombinant lectins in the characterization of a commercial biopharmaceutical product was investigated. The product was provided in two states, an optimally glycosylated product and an undesirable derivative of the product displaying immature glycans. A unsialylated derivative of the product was created by neuraminidase treatment. The binding profiles of PA-IL30, PL-IL30 and PA-IL_{mut1}30 to the product are shown in Fig 5.19. Having determined the specificities of each lectin in Section 5.5, they can be used to characterize the glycoprofile of the product.

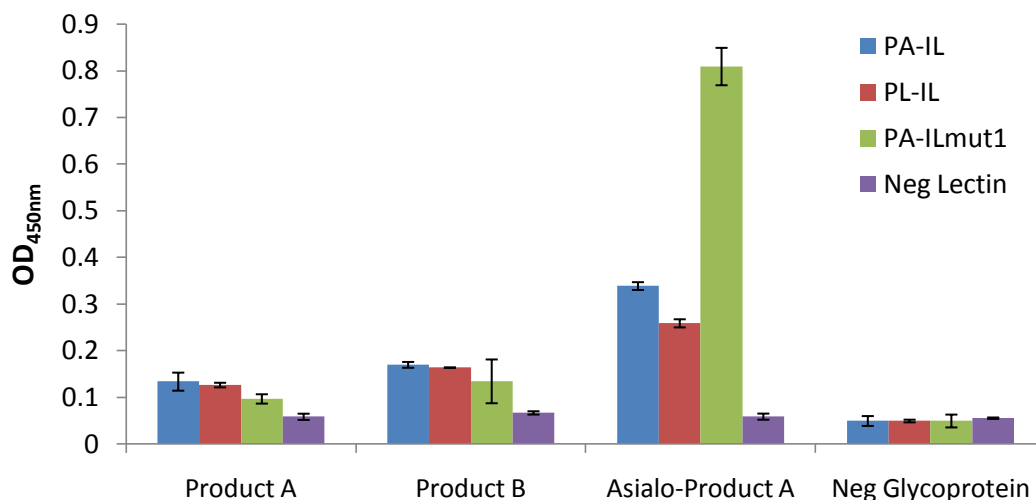


Fig 5.19 Binding of recombinant lectins to a biopharmaceutical product. Quantitative detection of recombinant lectins bound different versions of the same biopharmaceutical product assayed by the activity of an enzyme-linked secondary antibody (OD_{450nm}).

It has been shown in Section 5.4 that PA-IL displays a high specificity for α -galactose and a low specificity for GalNAc, that PL-IL broadly binds galactose molecules, and that PA-IL_{mut1} is strictly $\beta(1-4)$ specific. In Fig 5.19, the binding of these lectins to a biopharmaceutical product was investigated. Knowing the range of glycans each respective lectin binds to, it can be deduced from the assay that α -galactose is not present on either Product A or Product B. PA-IL displays low level activity, but as shown in Fig 5.11, this lectin can display low affinity to other galactose molecules. PL-IL shows similar levels of binding to the biopharmaceutical products, however this molecule has a broader specificity, and no conclusions can be drawn from this pattern. The PA-ILmut1 binding pattern clearly demonstrates the predominant glycan present beneath sialic acid is $\beta(1-4)$ galactose, as the binding signal dramatically increases upon neuraminidase treatment. It is also clear that in Product A and Product B, unsialylated protein is not present in high quantities, as this would expose the residue $\beta(1-4)$ galactose. The significance of this result in terms of the pharmaceutical industry is summarised in Section 1.4, where it was outlined that sialic acid is a key glycan in terms of efficacy and function.

5.8 Lectin affinity chromatography

The principles of lectin affinity chromatography (LAC) are outlined in Section 1.8.3. In Section 5.4 it was shown the lectin PA-IL_{mut1} showed a high specificity for asialylated derivatives of many asialylated glycoproteins. In Section 1.4 the importance of the presence of sialic acid residues to the biological activity of these therapeutics was outlined, and as a result the removal of unsialylated contaminant product during down-stream processing could significantly enhance the potency and the value for the producer. Thus far, the recombinant lectins have only been used as tools for the analysis of a purified glycoprotein. Applying the molecules into a downstream purification platform would greatly enhance the value of the lectins, so an experiment was set up to create a PA-IL_{mut1} column for the purposes of LAC. The column matrix was composed of cyanogen bromide activated sepharose-4B (Fig 5.20), which would bind to surface exposed primary amines on the lectin. The position of these lysine residues on the PA-IL_{mut1} structure can be seen in (Fig 5.21). The process of lectin immobilisation is outlined in Section 2.27, and the corresponding OD_{280nm} readings and SDS-PAGE analysis can be seen in Fig 5.21 and Fig 5.22. Having successfully immobilised lectin onto the column an attempt of capturing asialofetuin is shown in Fig 5.23.

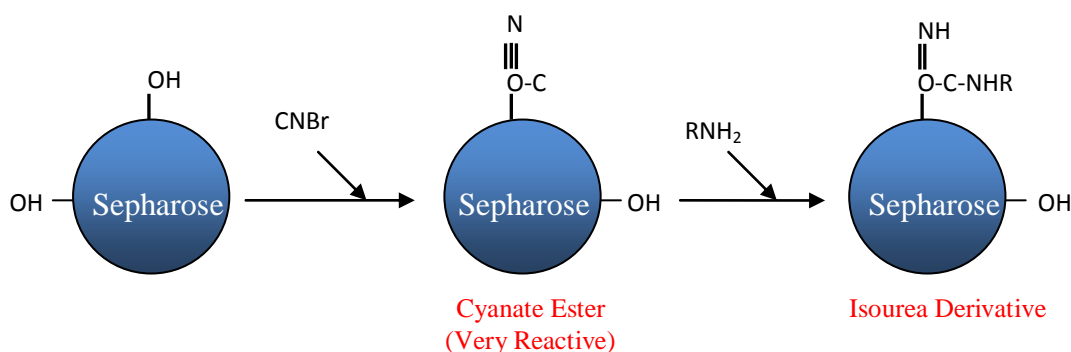


Fig 5.20 Activation of sepharose by cyanogen bromide and protein coupling to the activated matrix. Cyanogen bromide in basic conditions reacts with -OH groups on sepharose to form cyanate esters or imidocarbonates. These groups react readily with primary amines under very mild conditions; the net result is a covalent coupling of lectin, via lysine residues, to the agarose matrix.

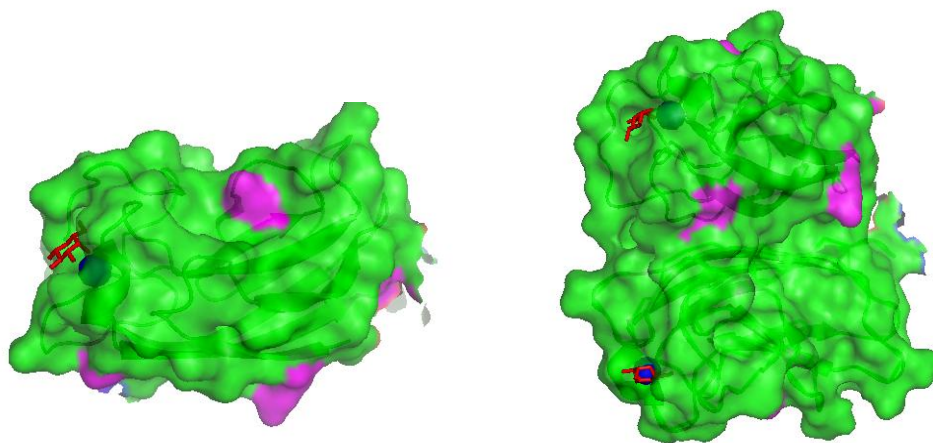


Fig 5.21 Location of surface exposed lysine residues on the PA-IL molecule. Cartoon representation of the location of lysine residues on the PA-IL molecule based on the crystal structure of Chen *et al.*, (1998). Lysine residues are highlighted in pink with the location of the sugar binding site indicated with a red galactose molecule. On the left is the monomeric structure, and the right is the proposed dimer. Image created using PyMol (Section 2.30).

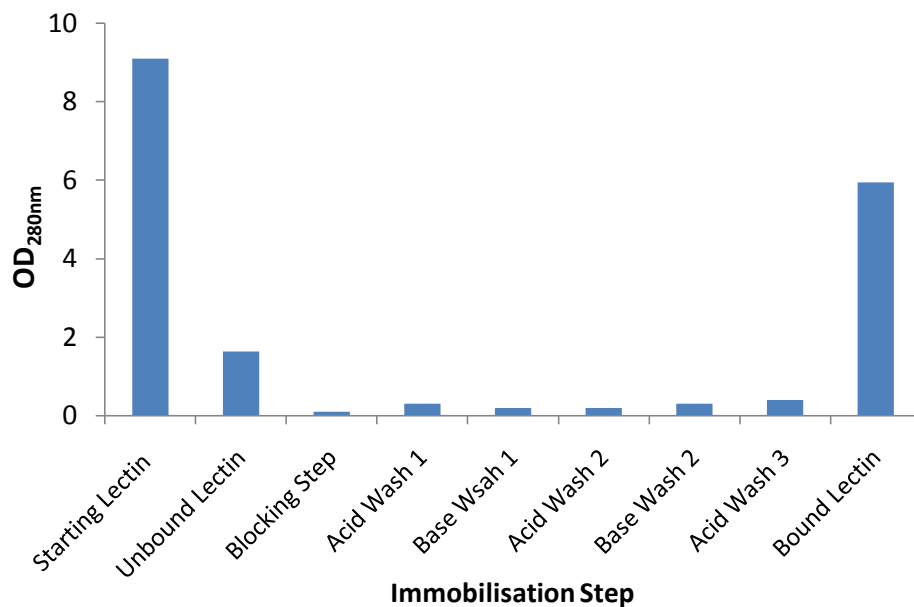


Fig 5.22 Determining the amount of PA-IL_{mut1} which successfully immobilised on activated sepharose. OD_{280nm} readings taken from 5 ml elutions during each step of the immobilisation process (Section 2. 27). The starting O.D of a solution of PA-IL_{mut1}30 was 8.96 which equates to a 5mg/ml solution. After mixing with the activated resin overnight, the flow-through an O.D of 1.72 was recorded, indicating that nearly 1 mg of protein did not immobilise. The subsequent wash steps show

minimal protein leaching. After subtracting the total amount of protein eluted from the protein from the initial concentration it was calculated that ~3 mg of lectin had been successfully immobilised on a 1 ml resin.

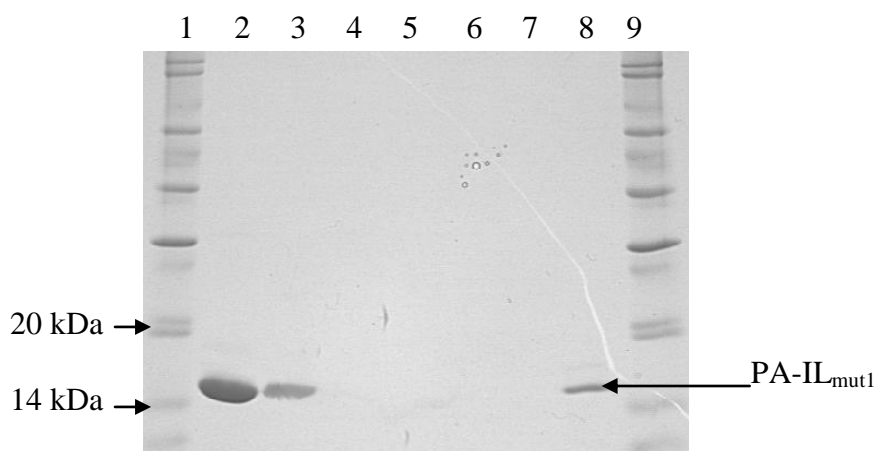


Fig 5.23 Immobilisation of PA-IL_{mut1} to activated sepharose. Analysis of the fractions from Fig 5.21 By 20% SDS-PAGE (Section 2.18). Lane 1, Molecular weight marker; Lane 2, Pa-IL_{mut1} in Immobilisation buffer; Lane 3, Unbound material; Lane 4, eluted blocking buffer, Lanes 5-7, acid/base wash steps; Lane 8, 10 μ l of boiled resin; Lane 9, Molecular weight marker.

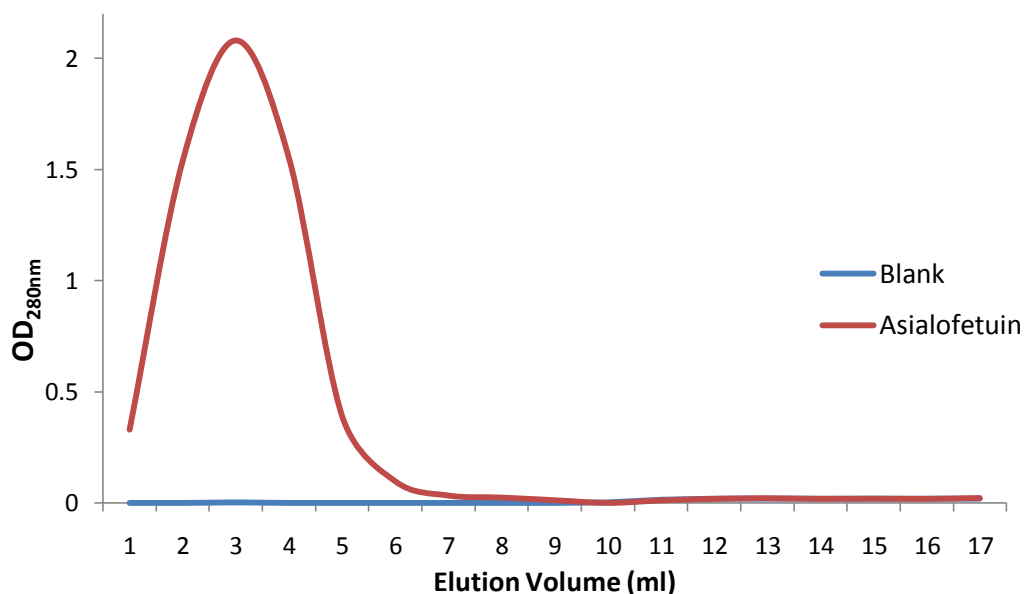


Fig 5.24 Application of asialothyroglobulin onto PA-IL_{mut1} column. OD_{280nm} values of successive 1 ml elutions from the PA-IL_{mut1} column following the application of asialothyroglobulin in binding buffer. At the 10 ml volume point, 10 mM galactose in binding buffer was added to the column.

There are four surface-exposed lysine residues per monomer on the PA-IL_{mut1} monomer (Fig 5.21). These were exploited in immobilisation of the molecule onto activated sepharose. Fig 5.22 displays how the amount of protein successfully immobilised was calculated. It was calculated that approximately 3 mg of lectin had been successfully immobilised, and upon SDS-PAGE analysis, it was discovered that a band of the correct size was present in a sample of boiled resin.

It was then attempted to use the constructed resin to purify asialothyroglobulin. It has already been shown in Section 5.5 that the lectin has a high affinity for this glycoprotein. Fig 5.23 shows the attempted purification. If the purification was a success, a peak at the 10ml point indicating eluted asialothyroglobulin would be expected. There proved to be no peak however, indicating that the PA-IL_{mut1}30 resin was inactive. Upon further investigation of Fig 5.21 it was clear that the most exposed lysine residue is K43, which is positioned close to the sugar binding site. It is possible that immobilisation via this amino acid would sterically hinder any binding of glycoproteins by the sugar binding site.

5.9 Discussion

The glycan binding profile seen for PA-IL is consistent (Table 5.1 and Fig 5.1) with results found in previous studies where structures containing terminal $\alpha(1-3)\text{Gal}$, $\alpha(1-4)\text{Gal}$ and $\alpha(1-6)\text{Gal}$ (Lanne *et al.*, 1994; Blanchard *et al.*, 2008) as well hydrophobic derivatives of galactose (Garber *et al.*, 1992) were found to bind wild-type PA-IL with the highest affinity. PA-IL has been shown previously to have a hemagglutination inhibition profile in the order phenyl βGal >melibiose>stachyose>raffinose>galactose, which is similar to the results obtained here (Chen *et al.*, 1998).

PL-IL on an amino acid level shows only 34% identity and 46% similarity to PA-IL (Section 1.9.2) yet, it has been shown that PL-IL displays similar binding profiles to its homologue PA-IL through both hemagglutination inhibition assays and ELLA. This is most evident from the similar inhibition profile seen in Table 5.1, and the same glycoproteins providing signal in Fig 5.7. Both proteins also have the same cation dependence (Fig 5.18), and prefer colder temperatures (Fig 5.2). Given the high homology between the two proteins especially within the sugar-binding loop this is not surprising.

However in terms of specific galactose containing oligosaccharides and glycoproteins, PL-IL displays binding properties with subtle but important differences to PA-IL (Fig 5.2). PL-IL displays a lower affinity for asialylated glycoproteins than PA-IL (Section 5.6), and low affinities to sepharose (Section 3.4.1) and phenyl--derivatives of galactose (Fig 5.2).

In terms of affinity values for the lectins, this study was limited in terms of available substrates. Oligosaccharides, though available, are very expensive, and the methodology for their immobilisation onto surfaces was not available within the laboratory. Similarly, neo-glycoconjugates (eg. BSA coated in specific oligosaccharide units) are extremely expensive, and glycan array technology is only available to groups that will publish the resulting data. This was not possible in this study due to extensive I.P. restraints. However an abundance of information can be procured from the limited resources that were available.

In the glycoprotein based ELLA, an image of what the lectin binds to in nature can be visualised, as free individual glycans are unlikely to be the intended targets for these proteins. Our results show that depending the glycoprotein in question, the PA-IL and PL-IL have similar binding profiles. However, PA-IL_{mut1} shows a different pattern (Figs 5.7A, 5.7B and 5.7C). The glycoproteins, all immobilised at the same level, that proved preferable to PA-IL and PL-IL were thyroglobulin asialo-thyroglobulin, asialo-hyaluronidase and asialo-ovalbumin. PA-IL_{mut1} was shown to bind to neuraminidase-treated versions of every glycoprotein screen with the exception of glucose oxidase (*A.niger*) and invertase (yeast), sources which would not be expected to encode for β -galactosyl transferases and hence the glycoproteins

are devoid of β -linked galactose residues. Results from glycan profiling studies of thyroglobulin, ovalbumin and hyaluronidase are limited but what they all possess is α -linked galactose moieties. For this reason, alpha-galactosidase was used to cleave off this residue in the case of asialothyroglobulin. The α -Gal specific commercial lectin used in this experiment was GSL-I, which did not show any binding to asialothyroglobulin, despite α -Gal residues being previously detected on the molecule (Rosenfeld *et al.*, 2007; Spiro *et al.*, 1984; Ronin *et al.*, 1986; Kamerling *et al.*, 1988). When treated with α -galactosidase, the binding signal of PA-IL and to a lesser extent PL-IL was reduced compared to the asialo-thyroglobulin signal, indicating some α -Gal binding, with no effect on PA-IL_{mut1} signal. β (1-4)galactosidase treatment had no effect on PA-IL binding, a very large effect on PA-IL_{mut1} binding, and a small effect on the PL-IL signal, indicating PA-IL_{mut1} to show a high specificity for β (1-4)galactose, which is abundant on the molecule. The commercial lectin specific for this residue is ECL, which had a reduction in signal after β (1-4)galactosidase treatment similar to the PA-IL_{mut1} reduction.

In the case of the glycoprotein fetuin, there have been no reports of α -galactose on the molecule, which would explain the low binding signals of PA-IL and PL-IL (Fig 5.10), but after the removal of sialic acid, the predominant terminal sugar is β (1-4)galactose. After β (1-4)galactosidase treatment, the lectins ECL and PA-IL_{mut1} both show a similar significant reduction of signal, indicating that both are binding to the same glycans. The reduction of signals by using galactosidase, particularly in thyroglobulin, are not completely back to baseline, but it should be considered that these enzymes are more unstable and less widely used than neuraminidases, and extremely expensive, hence optimisation of their activity proves difficult. Though PA-IL and PL-IL have no α -galactose on fetuin to bind to, there is clearly some binding occurring, which is reduced through the use of alpha-galactosidase rather than β -galactosidase. Both lectins have been shown to have low affinity for GalNAc (Fig 5.1), which is present on asialofetuin,

Coupling this data with the inhibition data seen in Fig 5.14, Fig 5.14 and Fig 5.16, where PA-IL was shown to be very quickly inhibited with α -Gal containing sugars, compared to PL-IL, which is not inhibited as efficiently by any of the sugars

investigated, it is clear that PA-IL and PL-IL, though binding to glycoproteins with similar affinities, have different specificities.

In summary, the highly characterised PA-IL molecule has had its specificity changed from $\alpha(1-3)\text{Gal}$ and $\alpha(1-4)\text{Gal}$ to $\beta(1-4)\text{Gal}$ by the mutation of three amino acids within the sugar binding site. The significance of this mutation becomes apparent in Section 5.7. Here a purified biopharmaceutical product is examined by PA-IL and PA-IL_{mut1} for $\alpha\text{-Gal}$ and $\beta(1-4)\text{Gal}$ residues. The $\alpha(1-3)\text{Gal}$ epitope is highly immunogenic, and would be targeted by the human innate immune system, with an estimated 1% of the total IgG molecules circulating in human serum targeting the sugar (Galili *et al*, 1984). The $\beta(1-4)\text{Gal}$ residue is an indicator of an unsialylated product, as it is exposed upon neuraminidase treatment. In a fermentation environment, neuraminidases are often released from animal cells at cell death, and detection of PA-IL_{mut1} binding could be a useful indicator of neuraminidase cleavage of sialylated products. Currently, the plant lectin most commonly used for this purpose is ECL. It has been attempted to produce a recombinant version of this lectin in *E. coli*, but it was shown to be insoluble, with purification from solubilised inclusion bodies required to obtain a functional molecule (Stancombe *et al.*, 2003). We propose that PA-IL_{mut1} would be a cheaper, and more functional alternative to ECL

6.0 Determination of the role of PL-IL in the life-cycle of *P. luminescens*

6.1 Introduction

The role of PL-IL in the life-cycle of *P. luminescens* in terms of its pathogenicity and its symbiotic relationship with *Heterorhabditis* nematodes was examined. The complex life-cycle of the bacteria and the nematode symbiont are summarized in Sections 6.2 and 6.3, with some facets of the life-cycle that could prove significant to the role of a potential adhesin such as PL-IL summarized in Section 6.4. As already shown in Chapter 5, PL-IL binds to alpha-linked galactose residues. It is also known to exist as a tetramer in its native form (Chapter 4). In this respect PL-IL shares its primary physical and affinity characteristics with the lectin PA-IL from *Pseudomonas aeruginosa*.

It was investigated whether PL-IL shares any functional characteristics with the homologue PA-IL, which is known to play an important role in the adhesion of *P. aeruginosa* to epithelial cell walls, and also to be involved in biofilm formation. This was done through a variety of established bio-assays carried out in the laboratory of Dr. David Clarke in U.C.C.

For these experiments, a mutant strain that lacked the lectin PL-IL was created by the insertion of a kanamycin resistance cassette within the gene *plu2096* that encodes the lectin, which in effect creates a mutant strain of *P. luminescens* which lacks the ability to express the lectin.

6.2 The bacterium *Photorhabdus luminescens*.

The bacterium *Photorhabdus luminescens* was initially described as *Xenorhabdus luminescens*. Its main characteristics include the ability to bioluminesce and to form symbiotic relationship with entomopathogenic nematodes (Thomas and Poinar, 1979). On the basis of phenotypic characteristics a new genus *Photorhabdus* was created (Boemare *et al.*, 1993). It is a member of the *Enterobacteriaceae* family and, therefore, is closely related to a number of common pathogens such as *Escherichia coli*, *Salmonella enterica* and *Yersinia pestis*. It forms mutualistic associations with specific entomopathogenic nematodes of the *Heterorhabditidae* (Fischer-Le Saux *et al.*, 1999). The two organisms have a complex life-cycle (Fig

1.11) that involves a symbiotic phase where the bacteria colonize the gut of the non-feeding infective stage nematodes (IJs). After the nematodes locate a susceptible insect host they release the facultatively anaerobic bacteria into the insect's hemolymph, where the bacteria enter a pathogenic stage of their life-cycle. *Heterorhabditis* enter the insect through respiratory spiracles, the mouth or the anus and migrate to the hemolymph. *P. luminescens* then in conjunction with the nematodes then induces septicaemia in the insect within 72 hours of infection, during which the nematode develops into a self-fertile hermaphrodite. Nematode development usually undergoes 3-4 generations, whereupon they develop once more into infective juveniles (IJs). These IJs are colonised by *P. luminescens* before the nematode emerges from the insect cadaver. Under laboratory conditions this takes approximately 14 days, and one initial IJ can result in the development of ~ 100,000 IJs (Forst *et al.*, 1997; Ciche *et al.*, 2006).

As a result of this multifaceted life-cycle, the bacteria must carry an array of proteins and secondary metabolites that can cater for mutualistic and pathogenic stages of growth. This range of proteins encompasses insecticidal toxins that lead to mortality in the insect, as well as antimicrobial compounds to repel competing micro-organisms (Akhurst, 1982; Richardson *et al.*, 1988; Williams *et al.*, 2005b), nematocides (Hu *et al.*, 1999), and factors that will allow for acquisition and symbiosis with the nematode host.

The genus is characterised by the ability to bioluminesce (through luciferase), and are the only terrestrial bacteria known to produce light. The genus has been divided into three subspecies; *Photorhabdus luminescens*, *Photorhabdus temperata* and *Photorhabdus asymbiotica*, the third of which comprises isolates from human wounds and was thought not to have a natural association with nematodes (Fischer-Le Saux *et al.*, 1999; Gerrard, 2003). Recently however, a nematode symbiont of *P. asymbiotica* has since been identified in the *Heterorhabditidae* genus (Gerrard *et al.*, 2006).

The genus is also characterised by the capacity to produce phase variants, which are designated primary and secondary forms (Smigielski *et al.*, 1994). There are some basic biochemical and morphological tests to distinguish between the two, as a stark

variance in the production of compounds such as antibiotics, pigments, and extracellular proteases exists between the two forms. Both forms are equally pathogenic when injected into the insect hemolymph, though bacteria recovered from the nematode are predominantly in the primary form. It is known that the primary form of *Photorhabdus* is required for nematode development and reproduction, and little is known about the role of the secondary form in nature, with some commentators suggesting it is an artefact of laboratory conditions, as it has never been seen in nature.

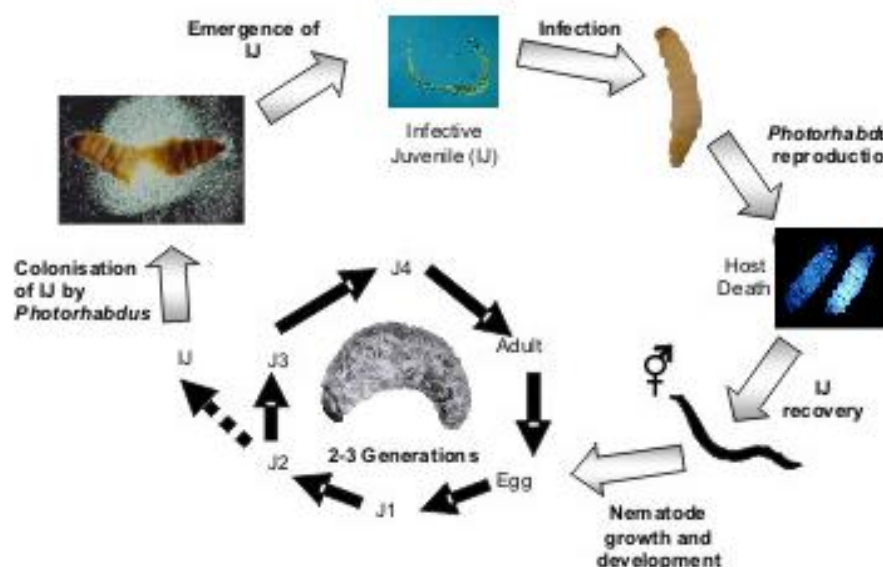


Fig 6.1 The life cycle of *Photorhabdus luminescens* and *Heterorhabditis bacteriophora* (Adapted from Ffrench-constant *et al.*, 2003).

6.2.1 Mutualism in *P. luminescens*

Microbial symbioses are ubiquitous in nature, with all plants and animals at some level associated with micro-organisms. One model system currently being used to understand these mutually beneficial relationships is the association of microbes from the genus *Photorhabdus* with the insect pathogen *Heterorhabditis*, which has a dependence on the bacterium for normal nematode growth and development (Joyce *et al.*, 2006).

The symbiotic relationship between *P. luminescens* and the nematode host can be divided into three distinct categories (Ffrench-Constant *et al.*, 2003). In the first stage, the bacteria exist in the lumen of the gut of the IJ nematodes, until entry of the nematode into an insect host when regurgitation of the bacteria into the hemolymph occurs. During this stage the bacteria produce signals that control nematode development and the recovery of adult *Heterorhabditis* as hermaphrodites (Strauch and Ehlers, 1998). Pathogenicity is tightly regulated at this stage also and the expression of virulence factors such as insecticidal toxin complexes and metalloproteases are tightly linked to the bacterial growth rate (Daborn *et al.*, 2001). In the next stage, the bacteria have grown to a high cell density and have entered the stationary phase of growth. In this stage a wide array of toxins and exoenzymes are released from the bacteria, for the purpose of inducing lethal septicaemia in the insect, and the bioconversion of the organic matter, which is in turn utilised by the nematode (Bowen *et al.*, 1998; Han and Ehlers, 2000). This relationship is very inflexible, as related organisms are often incompatible and cannot be substituted, with *P. luminescens* found to be pathogenic to the nematodes *Meloida incognita* (Hu and Webster, 1999), *Steinorma carpocapse* (Sicard *et al.*, 2004) and *Caenorhabditis elegans* (Sicard *et al.*, 2007). To date, the only gene that has been found to be specific for mutualism is *cip* (Bintrim and Ensign, 1998).

The genome of *P. luminescens* has been sequenced (Duchaud *et al.*, 2003), and a number of genes were identified with homology to virulence factors characterised in other pathogens (Ffrench-Constant *et al.* 2000). Some toxins bear no resemblance to known proteins, an example of which is Mcf-1 (makes caterpillars floppy), a high molecular weight toxin that confers on *E. coli* the ability to persist in, and kill an insect (Daborn *et al.*, 2002). This toxin appears to cause apoptosis in both the insect hemocytes and the midgut epithelium, a domain which is thought to be related to the BH-3 domain present in the protein, a feature found in many pro-apoptotic proteins (Baskin-Bey and Gores, 2005). Growth within the insect is essential for pathogenesis, and knock-out studies have to date identified genes required for *in-vivo* growth as being essential for virulence (Watson *et al.*, 2005). To date, there has been no investigation into the role of PL-IL in *Photorhabdus* virulence.

The final stage in the interaction with the nematode is the uptake of bacteria by the final generation of IJs. At this moment, the bacteria are still present in high densities, and thus far, an undiscovered colonization strategy commences which requires some specific interaction between the bacteria and the nematode gut, as unrelated bacteria have been found not to be retained.

To date, there has been no investigation into the role of PL-IL in the nematode-bacterial symbiosis.

6.2.2 Glycan profile of *P. luminescens* associated organisms

The bacterium *P. luminescens* is heavily dependant on the organisms *Heterorhabditis* as well as their insect prey, both of which have been the subjects of limited glycoprofiling studies. Insect cell lines are often seen as the intermediary between animal cells and yeast with respect to the degree of glycosylation seen on their respective glycoproteins (Lopez *et al.*, 1999). Whereas yeast cells predominantly decorate their proteins with mannose using the precursor structure Man₈GlcNAc₂-N-Asn, animal cells use glycosidase enzymes to cleave off the majority of mannose and add residues such as galactose, fucose and sialic acid using glycosyl-transferases (Section 1.2). Insects have many of these glycosidases, however they lack many of the glycosyl-transferases, and as a result the main structure present on insect glycoproteins is Man₃GlcNAc₂-N-Asn (Hollister *et al.*, 2002). When O-linkages are considered, the glycome becomes more complicated, with several reports of Gal and Gal-GalNAc being presented in this manner (Chen *et al.*, 1991).

The nematode *H. bacteriophora*, the other essential organism in the life-cycle of *P. luminescens*, like all other eukaryotes, shares the same initial steps in the process of N-glycosylation (i.e. the same core structure shown in Fig 1.1 is transferred to the protein within the ER). Much of characterization of nematode glycosylation has been done in *Caenorhabditis elegans*, which is closely related to *H. bacteriophora*.

C. elegans is found to display, like insect cells, high mannose containing glycans, as more complex glycans (i.e. those containing the sugars GlcNAc, Gal, and Neu5Ac) are uncommon. *C. elegans*, like *H. bacteriophora* goes through several stages of development, including a Dauer stage, which is very similar to the IJ stage (the stage of *P. luminescens* uptake and colonization by *H. bacteriophora*). The glycoprofile of this Dauer stage of development corresponds to more complex glycans than that of the other stages (Cipollo *et al.*, 2005), with some of these glycans being absent in any other stage of growth.

6.3 Construction of the *plu2096::kan^R* mutant

Gene specific cassette mutagenesis by allelic replacement is a standard procedure for the construction of knock out mutations in bacteria. This is facilitated by the suicide vector pJQ200sk+, which allows for the integration of mutated DNA fragments into a micro-organisms genome through homologous recombination (Quandt and Hynes, 1993), and the positive selection of a second recombination event which leaves the cassette intact in the genome. The positive selection is facilitated by the *sacB* gene which causes a suicide effect in the presence of sucrose. Therefore, once the plasmid is integrated into the genome of the bacteria by an initial homologous recombination event, selection on sucrose selects for cells in which the plasmid has been lost following a second recombination event in which the mutated copy of the gene is retained in the genome (Outlined in Fig 6.3).

To mutate the organism *P. luminescens*, vector constructs using the vector pJQ200sk+ are made in *E. coli*, and delivered by conjugation using a *mob* based system. Inserted into the pJQ200sk+ vector is the gene to be mutated as well as 1 kb of DNA that flanks that area of interest. In the centre of the gene of interest an antibiotic resistance cassette is inserted in the opposite orientation to the gene. A number of recombination events described in Fig. 6.3, allow for the elimination of

the pJQ200sk+ vector from the host cell, while incorporating the mutated gene into its own chromosome.

The *plu2096:kan^R* mutant was created in a *rif^R* *P. luminescens* TT01 strain. A 2kb region of DNA, which encompassed the gene *plu2096* and the sequence flanking on either side, was amplified in two sections, as an *Xba*I-*Pst*I and a *Pst*I-*Xho*I, which were inserted into the *Xba*I-*Xho*I sites within the MCS of the suicide vector pJQ200sk+. The *Pst*I restriction site created in the centre of the gene, served as the site for the insertion of the kanamycin resistance cassette from the plasmid pUC4K (Fig 6.2). An *Xho*I restriction site within the kanamycin cassette allowed for determination of the orientation of the cassette. This was incorporated into the genome by homologous recombination, and a successful clone validated by PCR (Fig 6.4).

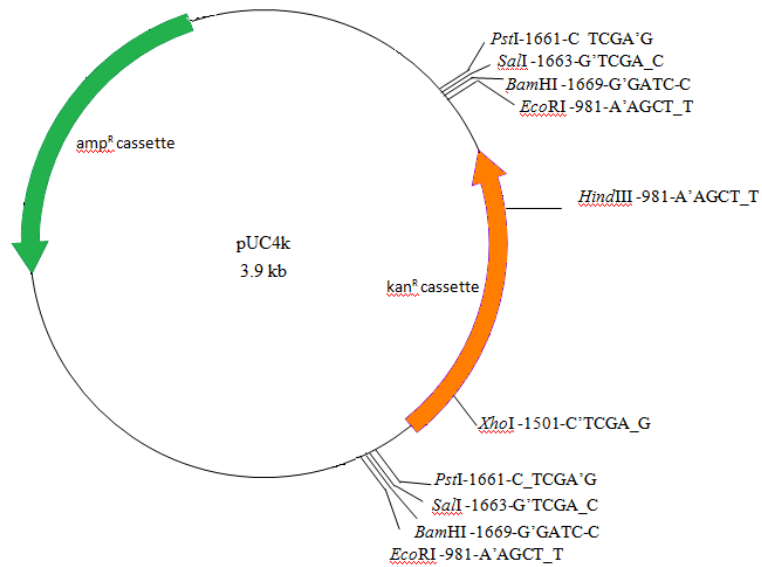


Fig 6.2 The pUC4K vector. The 3914 bp vector pUC4K (Table 2.2). The kanamycin resistance gene (orange) is flanked by multiple cloning sites to allow for the extraction of the cassette. An ampicillin resistance gene (green) is also present on the plasmid, but it wasn't used in this study.

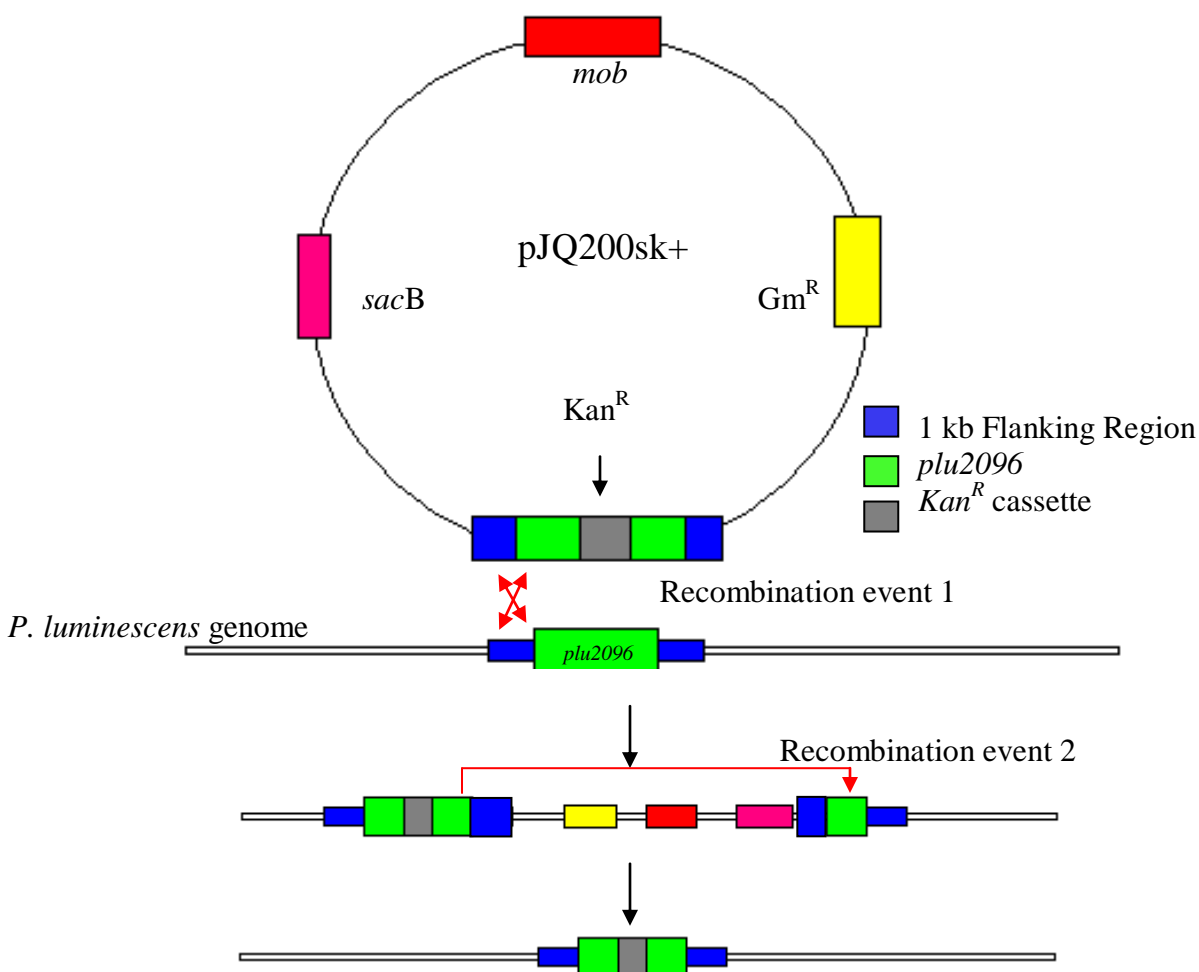


Fig 6.3 Schematic of homologous recombination events. The suicide vector pJQ200sk+ (Quandt and Hynes, 1993), containing the gene to be mutated, is shown with the following regions highlighted; the gene (green), the antibiotic resistance cassette (grey), 1 kb of chromosomal DNA flanking the gene (blue), gentamycin resistance (yellow), *sacB* gene (pink) and the mobilisation site (red). The first recombination event involves the incorporation of the entire vector into the chromosome through the interaction of regions bearing high homology (blue). The second recombination event involves the loss of the vector through internal recombination events.

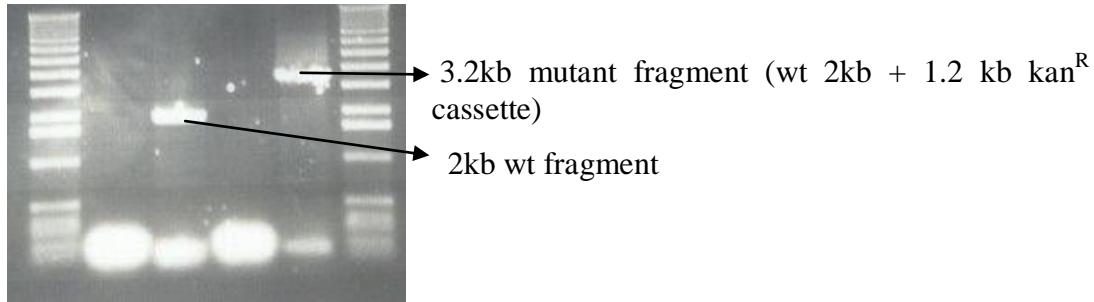


Fig 6.4 Confirmation of the *plu2096::kan^R* mutant by PCR. Agarose gel analysis of a PCR of wild type and mutant *P. luminescens* genomic DNA using the primers 2096KO-f and 2096KO-r (Table 2.3) which anneal 20 bp outside of the region amplified in the original cloning strategy.

6.4 The role of PL-IL in *P. luminescens* pathogenicity

The role of the lectin PL-IL in pathogenicity was determined by the pathogenicity assay (Section 2.33). In summary, *Galleria mellonella* larvae (supplied by Livefood, Rooks Bridge, Somerset, U.K.) were injected with a defined number of planktonic bacteria, with the time taken for insect death measured. This insect, commonly known as the greater wax moth, is a commonly used organism in *in-vivo* toxicology and pathogenicity modelling, as it has a well characterised immune response, and is an inexpensive and easy to handle insect.

In Fig 6.5 the comparison of wild-type and mutant pathogenicity is shown. This is typically displayed in terms of the LD₅₀ (the time at which 50% of the injected insects have died). There was little difference between the pathogenicity of the wild-type and the mutant, 48.5 and 48 hours respectively. After insect death, the cadaver turns a red colour and luminescence can be detected, as shown in Fig 6.7, as has been reported in similar studies (Silva *et al.*, 2002).

When both wild type and mutant bacteria were injected into the insect in an equal 50:50 ratio, similar quantities of each were re-isolated from the insect cadaver 48 hours later (Fig 6.6). This would indicate no competitive advantage in retention of the *plu2096* gene in terms of pathogenicity towards *Galleria* larvae.

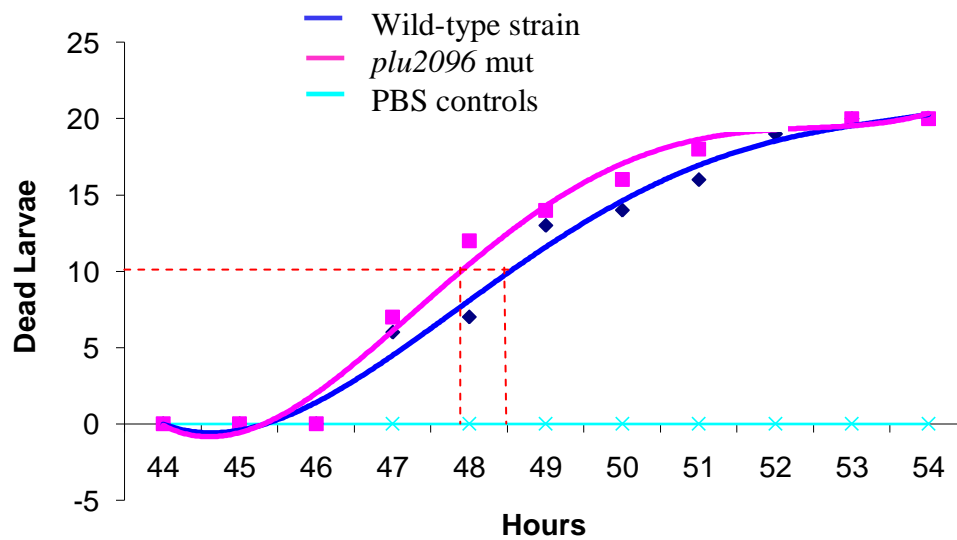


Fig 6.5 Pathogenicity of *P. luminescens* strains towards *Galleria* larvae. Twenty larvae were injected with 100 cells from each strain to assay for their pathogenicity. The LD₅₀ for each strain is marked with a dashed red line, with the blue and pink lines indicating the pathogenicity of wild-type and mutant trains respectively. The death rate of insects injected with PBS is marked in light blue.

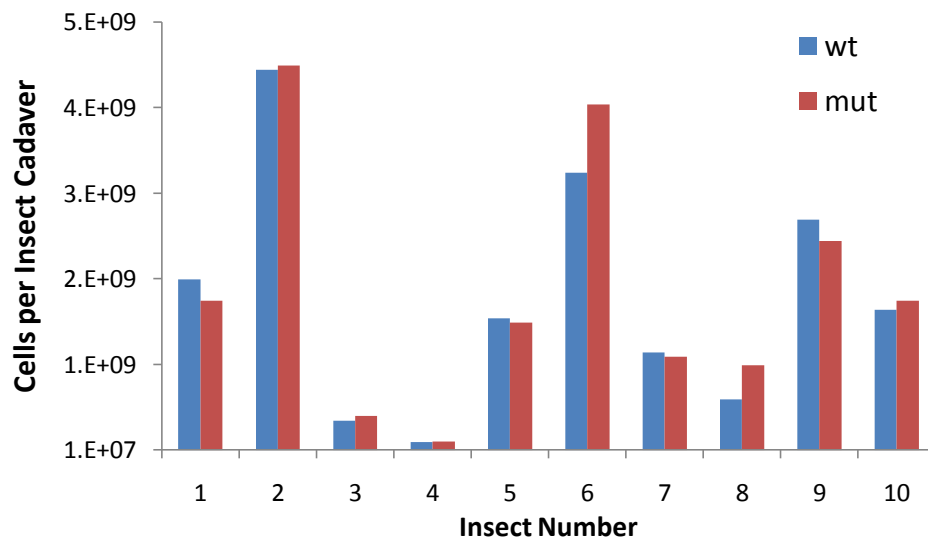


Fig 6.6 Comparison of *P. luminescens* wild-type and the mutant *plu2096::kan^R* in a competitive pathogenicity assay. The total number of each phenotype recovered from the cadaver of an insect that has been injected with a 50:50 mixture of both wild-type and *plu2096::kan^R* is depicted.

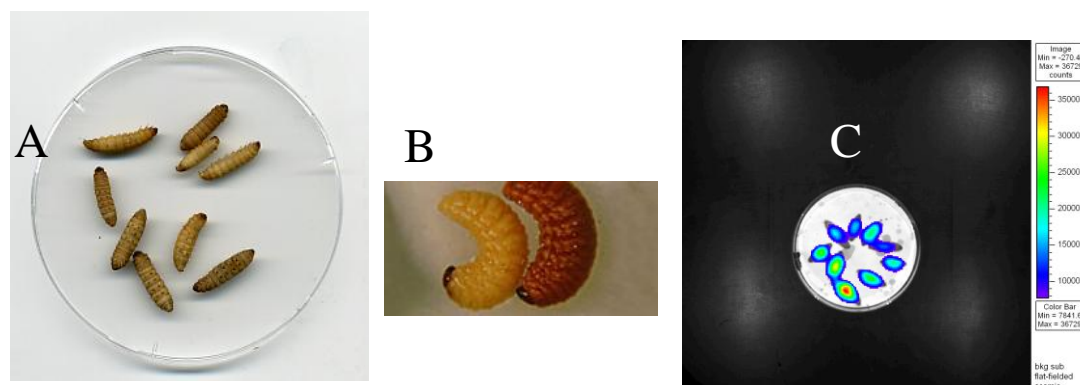


Fig 6.7 Infection of *G. mellonella* larvae by *P. luminescens*. Photographs showing *G. mellonella* larvae before and after infection with *P. luminescens*. Uninfected larvae are shown in Fig A, uninfected and infected larvae directly compared in Fig B, and the light emitted by the bacteria in infected larvae 48 hours post-infection is shown in C. (The image depicted in C was taken using an Andor fluorescence camera, 0.2 sec exposure time, f number 16. The amount of light emitted is measured in relative luminescence units (RLU), with a red colour indicating 35,000 and dark blue depicting less than 10,000).

6.5 The role of PL-IL in *P. luminescens* symbiosis

The symbiosis assay (Section 2.34) determines if *P. luminescens* is capable of colonizing the nematode. The IJ nematodes are added to a plate overlain with the appropriate bacteria, whereupon they mature through the early stages of their life-cycle. At the J4 stage (See Fig 6.1) they take up the bacteria, which will then colonize the gut of the nematode. The nematode matures into an adult, from which all daughter nematodes will not only contain identical genetic material, but also the same bacteria that were present in the parent nematode.

When nematodes are harvested from plates inoculated with either the wild-type or the mutant strain, the bacteria that are harvested from their guts were found to have the kanamycin resistance phenotypes shown in Fig 6.8. The wild-type strain is kanamycin sensitive, while the mutant is kanamycin resistant as a result of the kanamycin resistance cassette inserted in the *plu2096* gene.

From the symbiosis assay it was shown that the mutation of *plu2096* did not result in any loss in the ability to colonize the nematodes. The numbers of nematodes recovered from wild-type and mutant inoculated plates were both in the region of 10,000. To determine if the mutation resulted in reduced colonisation ability by comparison with the wild-type, a competitive symbiosis assay was set up.

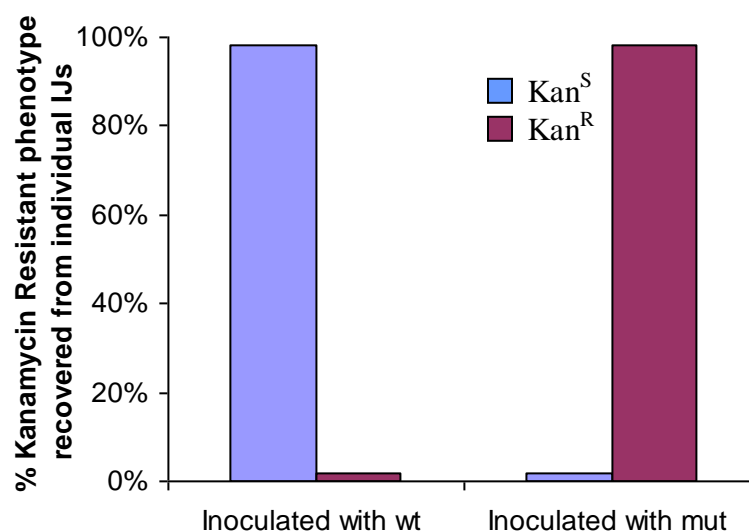


Fig 6.8 Colonization of *Heterorhabditis* nematodes by wild-type and mutant strains. *Heterorhabditis* nematodes were added to plates containing only the wild-type (Kan^S) or the mutant (Kan^R) strains. In each case the inoculated bacteria were recovered from the next generation of nematodes and the kanamycin resistance phenotype was determined.

6.5.1 Competitive symbiosis assay

To set up a competitive symbiosis assay, an equal proportion of wild-type and mutant strains are spread on lipid agar plates. Nematodes are added to the plates once a mat of bacteria has developed and any advantage of one strain over the other to colonize the nematodes should be identifiable from the colonization pattern of the next generation of nematodes. As controls, the proportion of both strains present on the plate was estimated at point of inoculation, and at 3 and 24 days. This served to determine that the strains were equally represented when colonisation of the

nematodes was initiated at day 3. The counts for days 3 and 24 also enabled a comparison to be made between the strains with respect to their survival in the absence of nematodes.

Bacterial samples were taken from the plates at three time periods, diluted in PBS and plated on medium in the presence and absence of kanamycin to estimate the relative abundance of the wild type and mutant bacteria.

The ratio of wild-type and mutant bacteria was determined at day zero to be 50:50 (Fig 6.8). At day three, within error, both strains retained the same ratio, however by day 24 the wild-type had outgrown the mutant by a factor of 10%. Day 3 corresponds to the time that nematodes were added to a separate set of plates on which symbiosis assays were conducted.

After 24 days bacteria were isolated from 50 nematodes and the proportion of Kan^s and Kan^r isolates was determined by plating on medium containing kanamycin. Fig 6.11 shows the percentage of nematodes that were found to have been colonised by each strain alone, by the two strains together and by neither strain.

Ten individual plate assays were conducted to measure competition for colonisation and symbiosis. From each plate the relative proportion of mutant and wild type was determined and compared with the proportions recovered from nematodes that were grown on those plates. The results from four of these plates are shown in Fig 6.10. It is notable that in one case, although the wild type was most predominant among the free living bacteria the mutant had colonised all three of the nematodes that were screened.

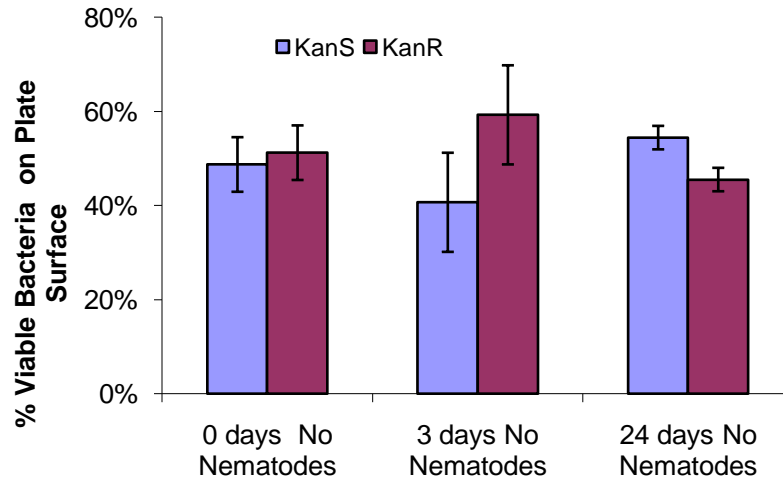


Fig 6.9 Proportions of strains on lipid agar plates in the absence of nematodes. The percentage of each phenotype growing on control plates at set time-points for the competitive symbiosis assay.

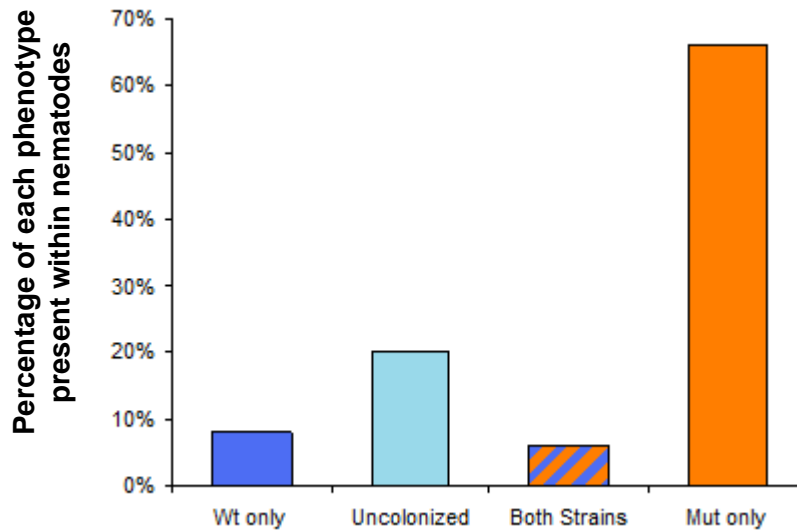


Fig 6.10 Proportion of each bacteria type present in the colonized nematodes in the competition assay at day 24. The percentage of each bacterial phenotype present within crushed nematodes at the end of the competitive symbiosis assay.

6.6 Discussion

From the literature it is known that the specific sugar residues for which PL-IL displays a high affinity (α -Gal) are not present to a large extent on cell surfaces of glycoproteins within the majority of insect cell lines (due to the fact that insects lack many of the relevant galactosyl-transferases) (Chen *et al.*, 1991; Hollister *et al.*, 2002). As a result it was not unexpected that the *plu2096* mutant would not have reduced pathogenicity towards the *Galleria mellonella* larvae (Fig 6.5). This result was confirmed by a competition pathogenicity assay (Fig 6.6), where it was found that neither strain of the bacteria out-grew the other within a cadaver of a dead insect when injected into a live insect in equal proportions.

It can thus be said that the lectin isn't crucial for pathogenicity in *Galleria mellonella*, though it is worth noting that the activity of a lectin from the PA-IL family has previously been shown to be sensitive to temperature (Gilboa-Garber and Sudakevitz., 1999). All of the assays presented here were performed at 25°C, which is not the typical temperature at which these organisms are found in nature. After conducting these bio-assays, it was then found that PL-IL activity is very sensitive to temperature (Section 5.2), so it is possible that the molecule plays more of a role in pathogenicity under different conditions. Also, mutants of *P. luminescens* that have been shown to be non-pathogenic in *Galleria mellonella* have sometimes proved to be pathogenic in other insects such as *Manduca sexta*, and vice-versa, so the lectin could have a more important role for the organism's pathogenic life-cycle than proposed here.

In the symbiosis assay, the ability of nematodes to be colonized by *P. luminescens* and then reproduce is assayed, as germ-free nematodes do not reproduce as efficiently as their colonized counterparts (Han and Ehlers, 2000). It was found in these experiments that both the wild-type and the lectin mutant colonized the nematodes and allowed nematode reproduction (Fig 6.8). It could be concluded that the protein product of *plu2096* is not essential for nematode symbiosis.

In the competitive symbiosis assays, the crushing of 50 nematodes yielded the result that 66% of the juvenile nematodes contained mutant bacteria, compared with 8% that contained the wild-type only (Fig 6.10), which indicates a clear advantage, in terms of nematode colonization, for the bacteria that do not encode PL-IL. If the nematodes are not selectively taking up the mutant, then the prevalence of the mutant in the guts of the next generation of bacteria (compared to the wild-type) (Fig 6.10) must be explained. This result should only be interpreted in terms of the proportion of bacteria that were available to the nematodes at the time of colonization (Fig 6.9). As the experiment is dependent on there being an equal ration of both strains, any deviation from this can have a serious effect on the final colonisation data. As shown in Fig 6.9 at day 3, at the time of colonization, the strains were present on the lipid agar plates in equal proportion, within the boundaries of error. However, notwithstanding the error bars, the results in Fig 6.9 suggest that the mutant may have been present in slightly higher numbers at the time of colonisation.

It would also be of interest to investigate inoculation with weighted proportions e.g. where the colonisation with an inoculum weighted 10:1 in favour of the mutant or 10:1 in favour of the wild type would be compared with a 1:1 proportion. If the mutant is indeed more aggressive as a colonizer then this may emerge from weighted infection ratios.

A recent paper reported the conclusion that it is usually only one of the ingested bacteria that adheres to the nematode intestinal wall prior to invasion of the juveniles from the maternal nematode, in order to prevent ‘cheating’ cells that will not contribute to virulence (Ciche *et al.*, 2008). In the experiments reported here three nematodes from fifty were found to be colonized by both bacterial strains in disagreement with Ciche *et al.*, (2008), indicating some co-infection. As this molecule is present in significant amounts in the extracellular fraction of *P. luminescens* culture (Turlin *et al.*, 2006), it should be investigated whether the mutants have colonized more effectively by utilizing the protein that has been produced by other cells, and are in effect ‘cheating’. Comparison of the original

mutant with a mutant constructed by either a deletion of *plu2096*, or insertion of another cassette would ascertain if this was the case.

This study was an initial study investigating the importance of the lectin in the symbiotic relationship between *P. luminescens* and *Heterorhabditis bacteriophora*, with the conclusion that strains that lack this molecule are not any more defective in their ability to colonize the nematode gut than the wild-type. A more comprehensive study would have to evaluate how these mutants behave in the animal model, as this symbiosis assay removes the insect from the natural life-cycle, where this lectin could play a more critical role.

7.0 Conclusions and Recommendations

In this research thesis, the genes that encode *P. aeruginosa* PA-IL and *P. luminescens* PL-IL were cloned into bacterial expression systems, enabling high level expression and purification of the recombinant lectins. A novel lectin was then created which contained the binding site of PL-IL, as well as the core lectin skeleton of PA-IL (Chapter 3). Extensive physical characterisation of these recombinant lectins was then undertaken (Chapter 4) before the affinity and specificity of the molecules were investigated by a series of assays (Chapter 5). Finally, preliminary studies into the role of PL-IL in *P. luminescens* were then carried out (Chapter 6).

Through bioinformatics analysis, it was discovered that the gene *lecA*, which encodes PA-IL, has a number of homologues (Section 1.9.2), which had not previously been the subject of any characterisation studies. The data revealed in this thesis contributes significant insight into not only PA-IL, the original member of this PA-IL superfamily, but also its homologue PL-IL. By constructing a mutant lectin containing a novel sugar binding site, the study was further expanded to encompass a novel member of the family. It was theorized that in order to develop a larger range of lectins with different specificities, cloning the binding site from one homologue into the PA-IL molecule would eliminate the need to clone entire molecules from other strains. However from Section 5.5, it was seen that cloning a PL-IL binding site with a low specificity and affinity for α -galactose and β -galactose into PA-IL, did not recreate a molecule with the same specificity, but instead created a molecule with a high specificity for β -galactose with no specificity for α -galactose. It was therefore concluded that by the construction of mutants in PA-IL it would be possible to create novel binding specificities, and not only re-create existing specificities that exist within the family.

PA-IL and PL-IL, are the most distally related members of the PA-IL family as shown by bio-informatic analysis (Fig 3.38), and if a similar approach was to be taken again, PA-IL would not be the core skeletal molecule used, but rather PL-IL, as it is more closely related to the other members of the family. PL-IL is also well characterised as a result of this study, with detailed information on its multimeric assembly (Chapter 4) as well as temperature (Section 5.2) and co-factor requirements (Section 5.7).

The next step in the characterisation of the PL-IL molecule would be X-ray crystallography, as only limited information can be obtained from a predicted model based on sequence homology (Section 4.2). As a result of the expression optimisation carried out in Chapter 3, large amounts of soluble recombinant lectin could be purified from *E. coli* without having expended every possible optimisation avenue. Crystallisation requires a large amount of purified protein substrate, due to an extensive screening process that examines the compatibility of numerous different crystallisation buffers and conditions. The total yields for purified recombinant lectin per 500/ml culture are given in Table 3.3. For the purposes of this thesis, these yields were sufficient for all subsequent characterisation studies, but if higher yields of expression were desired, several options remain regarding further optimisation of expression. For example, plasmid stability was not investigated in this study, and poor plasmid maintenance within the cells can seriously affect protein expression levels. The introduction of a more stable antibiotic resistance cassette, such as carbenicillin, as ampicillin is well known to be highly unstable, is one such possible method. Another possible area for optimisation is media composition, as LB is the only medium used throughout this study, but several alternatives exist, for example terrific broth, SOB broth, or several proprietary enriched media.

The availability of a crystal structure would allow for further site-directed mutagenesis of the molecule, to elucidate which individual amino acids play the most important role in sugar binding. Of further interest would be a random mutagenesis study what would concentrate on specific sugar-binding residues. It has been shown in this thesis the ease with which PA-IL had its specificity changed from α -galactose to β -galactose, and that was incorporating residues seen in a related molecule. Through the incorporation of very different residues it could be possible to change the affinity of the molecule for other glycan residues such as mannose or sialic acid.

The PA-IL family of molecules has already been shown to provide robust and reproducible results in the ELLA assay (Section 5.4), which have been used in the characterisation of glycoprotein-based therapeutics (Section 5.6). With the

possibility of further additions to the family through site-directed mutagenesis, an array of specificities may be created within the basic framework of one lectin molecule. This could ultimately lead to the creation of a uniform recombinant lectin microarray with each well containing the same core lectin molecule, but each with a variety of different glycan specificities. The main benefits to such a lectin-array would be every molecule having the same optimal conditions (pH, temperature, co-factor requirements and buffer composition) and the ease at which these recombinant lectins can be produced to a high level of purity. Recombinant expression of bacterial lectins also allows for the introduction of immobilisation tags into the protein structure. Directed immobilisation via these tags would create a more homogenous distribution of lectin on the array surface.

The possibility of incorporation of the molecules into a lectin affinity chromatography platform was also briefly investigated. PA-IL_{mut1} (Section 5.8) was successfully immobilised on a sepharose matrix, but there was no indication of successful glycoprotein capture for the lectin. As a result of the extensive physical characterisation carried out in Chapter 4, the effect on lectin activity by the addition of affinity tags to either terminus is known. As a result of this a lysine tag could be incorporated onto the C-terminus of the lectins to allow for a more directed immobilisation approach, which it is hoped would result in a more active lectin-resin. Alternatively, poor glycoprotein capture could be the result of relatively low affinities for both molecules. Altering affinities can be achieved using the same site-directed mutagenesis approach that was used for changing specificities.

Finally, this thesis included an initial study investigating the importance of the lectin in the symbiotic relationship between *P. luminescens* and *Heterorhabditis bacteriophora*, with the conclusion that strains that lack this molecule are not any more defective in their ability to colonize the nematode gut than the wild-type. Mutant strains lacking the molecule were also found not to be any less pathogenic than the wild type strain within the *Galleria* insect model, indicating no role in pathogenesis. A more comprehensive study would have to evaluate how these mutants behave in the animal model, as the symbiosis assay removes the insect from

the natural life-cycle, while the pathogenesis assay removes the nematode from the natural life-cycle, where this lectin could play a more critical role.

In conclusion, this study, and novel lectin characterisation studies like this, will prove invaluable to the increasingly important field of glycobiology, as the biopharmaceutical industry will require a wide variety of molecules such as these recombinant lectins that will ultimately be used to obtain a comprehensive glycoprofile of the next generation of recombinant glycoprotein therapeutics.

8.0 References

- Adam, J., Pokornà, M., Sabin, C., Mitchell, E.P., Inmberty, A., and Wimmerová, M. (2007) Engineering of PA-III lectin from *Pseudomonas aeruginosa* – Unravelling the role of the specificity loop for sugar preference. *B.M.C. Struct. Biol.* **7**(36)
- Adar, R., Streicher, H., Rozenblatt, S., and Sharon, N. (1997) Synthesis of soybean agglutinin in bacterial and mammalian cells. *Eur J Biochem.* **249**:684–9.
- Aebi, M., and Hennet, T. (2001) Congenital disorders of glycosylation: Genetic models lead the way. *Trends in Cell Biol.* **11**(3):136-141
- Akhurst, R.J. (1982) Antibiotic activity of *Xenorhabdus spp.*, bacteria symbiotically associated with insect pathogenic nematodes of the family *Heterorhabditidae* and *Steinernematidae*. *J. Gen. Microbiol.* **128**(12):3061-3065
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**(17): 3389-402.
- Ambrosi, M., Cameron, N.R., and Davis, B.G. (2005). Lectins: tools for the molecular understanding of the glycode. *Organ. Biomol. Chem.* **3**(9): 1593-1608.
- Apweiler, R., Hermjacob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from the SWISS-PROT database. *Biochim. Biophys. Acta.* **1473**: 4-8
- Arora, S., Wolfgang, M.C., Lory, S., and Ramphal, R. (2003) Sequence polymorphism in the glycosylation island and glagellins of *Pseudomonas aeruginosa*. *J. Bacteriol.* **186**(7): 2115-2122
- Avichezer, D., Katcoff, D.J., Garber, N.C., and Gilboa-Garber, N. (1992). Analysis of the amino acid sequence of the *Pseudomonas aeruginosa* galactophilic PA-I lectin. *J. Biol. Chem.* **267**(32): 23023-23027.

- Bajolet-Laudinat, O., Girod-de Bentzmann, S., Tournier, J.M., Madoulet, C., Plotkowski, M.C., Chippaux, C., and Puchelle, E. (1994). Cytotoxicity of *Pseudomonas aeruginosa* internal lectin PA-I to respiratory epithelial cells in primary culture. *Infect. Immun.* **62**(10): 4481-7.
- Barkai-Golan, R., Mirelman, D., and Sharon, N. (1978) Studies on growth inhibition by lectins of *Penicillia* and *Aspergilli*. *Archives of Microbiol.* **116**(2): 119-124
- Baskin-Bey, E., and Bores, G. (2005). Death by association: BH3 domain only proteins and liver injury. *Am. J. Physiol. Gastrointest. Liver Physiol.* **289**: 987-990
- Bates, P.A., Kelley, L.A., MacCallum, R.M. and Sternberg, M.J.E. (2001) Enhancement of Protein Modelling by Human Intervention in Applying the Automatic Programs 3D-JIGSAW and 3D-PSSM. *Proteins: Structure, Function and Genetics, Suppl.* **5**:39-46
- Bewley, C. (2001) Solution structure of a cyanovirin-N:Man alpha 1-2Man alpha complex: structural basis for high affinity carbohydrate-mediated binding to gp120. *Structure.* **9**(10): 931-940
- Bintrim, S.B., and Ensign, J.C. (1998) Insertional inactivation of genes encoding crystalline inclusion bodies of *Photobacterium luminescens* results in mutants with pleiotropic phenotypes. *J. Bacteriol.* **180**(5): 1261-1269
- Birnboim, H.C., and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research.* **7**: 1513-1523
- Blanchard, B., Nurisso, A., Hollville, E., Tétaud, Wiels, J., Pokorná, Wimmerová, Varrot, A., and Imberty, A. (2008) Structural basis of the preferential binding for globo-series glycosphingolipids displayed by *Pseudomonas aeruginosa* Lectin I. *J. Mol. Biol.* **383**: 837-853
- Boemare, N.E., Akhurst, R.J., and Mourant, R.G. (1993) DNA relatedness between *Xenorhabdus* spp. (*Enterobacteriaceae*), symbiotic bacteria of entomopathogenic nematodes, and a proposal to transfer the *Xenorhabdus luminescens* to a new genus, *Photobacterium* gen. nov. *Int. J. Syst. Bacteriol.* **43**: 249-255

- Boteva, R.N., Bogoeva, V.P., and Stoitsova, S.R. (2005). PA-I lectin from *Pseudomonas aeruginosa* binds acyl homoserine lactones. *Biochim. Biophys. Acta.* **1747**(2): 143-9.
- Bowen, D., Rocheleau, T.A., Blackburn, M., Andreev, O., Golubeva, E., Bhartia, R., and Ffrench-Constant, R.H. (1998). Insecticidal toxins from the bacterium *Photobacterium luminescens*. *Science.* **280**(5372): 2129-2132.
- Boyd, W.C., and Shapleigh, E. (1954) Specific precipitating activity of plant agglutinins (Lectins). *Science.* **119**(3091):419
- Branco, A., Bernabè, R., Ferreira, B., de Oliveira, M., Garcia, A., and Filho, G. (2004). Expression and purification of the recombinant SALT lectin from rice (*Oryza sativa* L.). *Protein Expression and Purification* **33**: 34-38.
- Brockhausen, I. (2006) Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO reports.* **7** (6): 599-604
- Böhme, U., and Cross, G.A. (2002) Mutational analysis of the variant surface glycoprotein GPI-anchor signal sequence in *Trypanosoma brucei*. *J. Cell Sci.* **115**(4): 805-816
- Böhme, U., and Cross, G.A.M. (2002) Mutational analysis of the variant surface glycoprotein GPI-anchor signal sequence in *Trypanosoma brucei*. *J. Cell Sci.* **115**: 805-816
- Campbell, C.Y., and Yarema, K.J. (2005) Large scale approaches for glycobiology. *Genome Biol.* **6**(11): 236-244
- Cautrecasas, P., Wilchek, M., and Anfinsen, C.B. (1968) Selective enzyme purification by affinity chromatography. *Biochem.* **61**: 636-643
- Chao, Q., Casalongue, C., Quinn, J.M, Etzler, M.E. (1994) Expression and partial characterization of *Dolichos biflorus* seed lectin in *Escherichia coli*. *Arch. Biochem. Biophys.* **313**:346–50.
- Chen, C.P., Song, S.C., Gilboa-Garber, N., Chang, K.S., and Wu, A.M. (1998). Studies on the binding site of the galactose-specific agglutinin PA-IL from *Pseudomonas aeruginosa*. *Glycobiol.* **8**(1): 7-16.

- Chen, W., Shen, Q.X., and Bahl, O.P. (1991) Carbohydrate variant of the recombinant β -subunit of human choriogonadotropin expressed in baculovirus expression system. *J Biol Chem.* **266**(7):4081-4087
- Ciche, T.A., Darby, C., Ehlers, R., Forst, S., and Goodrich-Blair, H. (2006) Dangerous liaisons: the symbiosis of entmopathogenic nematodes and bacteria. *Biological Control.* **38**: 22-46
- Ciche, T.A., Kim, K.S., Kaufmann-Daszczuk, B., Nguyen, K.C., and Hall, D.H. (2008) Cell invasion and matricide during *Photorhabdus luminescens* transmission by *Heterorhabditis bacteriophora* nematodes. *Appl. Enviro. Microbiol.* **74**(8):2275-2287
- Cioci, G, Mitchell, E.P., Gautier, C., Wimmerová, M., Sudakevitz, D., Pèrez, S., Gilboa-Garber, N., and Imberty, A. (2003). Structural basis of calcium and galactose recognition by the lectin PA-IL of *Pseudomonas aeruginosa*. *FEBS Lett.* **555**(2): 297.
- Cipollo, J.F., Awad, J.M., Costello, C.E., and Hirschberg, C.B. (2005) N-Glycans of *Caenorhabditis elegans* are specific to development stages. *J. Biol. Chem.* **280**: 26063-26072
- Daborn, P. J., Waterfield, N, Silva, C.P., Au, C.P., Sharma, S., and Ffrench-Consant, R.H. (2002). A single *Photorhabdus* gene, makes caterpillars floppy (mcf), allows *Escherichia coli* to persist within and kill insects. *PNAS.* **99**(16): 10742-10747.
- Daborn, P.J., Waterfield, N., Blight, M.A., and Ffrench-Constant, R.H. (2001). Measuring virulence factor expression by the pathogenic bacterium *Photorhabdus luminescens* in culture and during insect infection. *J. Bacteriol.* **183**(20): 5834-5839.
- Danguy, A., Camby, I., and Kiss, R. (2002). Galectins and cancer. *Biochim. Biophys. Acta.* **1572**(2-3): 285-93.
- Dell, A., and Morris, H.R. (2001) Glycoprotein structure determination by mass spectrometry. *Science.* **291**:2351–2356

- Diggle, S., Winzer, K., Lazdunski, A., Williams, P., and Cámara, M. (2002). Advancing the quorum in *Pseudomonas aeruginosa*: MvaT and the regulation of *N*-Acylhomoserine lactone production and virulence gene expression. *J. Bacteriol.* **184**(10): 2567-2586.
- Diggle, S.P., Stacey, R.E., Dodd, C., Cámara, M., Williams, P., and Winzer, K. (2006). The galactophilic lectin, LecA, contributes to biofilm development in *Pseudomonas aeruginosa*. *Environ. Microbiol.* **8**(6): 1095-1104.
- Dosanjh, A., Lencer, W., Brown, D., Ausiello D.A., and Stow J.,L. (1994) Heterologous expression of of delta F508 CFTR results in decreased sialylation of membrane glycoconjugates. *Am. J. Physiol. Cell Physiol.* **266**: 360-366
- Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givauden, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M., Charles, J., Dassa, E., Derose, R., Derzelle, S., Freysennet, G., Gaudriault, S., Médigue, C, Lanois, A., Powell, K., Siguier, P., Vincent, R., Wingate, V., Zouine, M., Glaser, P., Boemare, N., Danchin, A., and Kunst, F. (2003). The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotechnol.* **21**(11): 1307-13.
- Dwek, M. V., Ross, H., and Leathern, A. (2001). Proteome and glycosylation mapping identifies post-translational modifications associated with aggressive breast cancer. *Proteomics.* **1**(6): 756-62.
- Eisen, M.B., Sabesan, S., Skehel, J.J., and Wiley, D.J. (1997). Binding of the influenza A virus to cell-surface receptors: Structures of five hemagglutinin-sialyloligosaccharide complexes determined by x-ray crystallography. *Virol.* **232**(1): 19-31.
- Elliott, S., Lorenzini, T., Asher, S., Aoki, K., Brankow, D., Buck, L., Busse, L., Chang, D., Fuller, J., Grant, J., Hernday, N., Hokum, M., Hu, S., Knudten, A., Levin, N., Komorowski, R., Martin, F., Navarro, R., Osslund, T., Rogers, G., Rogers, N., Trail, G., and Egrie, J. (2003) Enhancement of therapeutic protein *in vivo* activities through glycoengineering. *Nat. Biot.* **21**(4): 414-421

- Engvall, E. and P. Perlman (1971). Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochem.* **8**(9): 871-4.
- Erickson, B., Deyde, V., Sanchez, A., Vincent, M., and Nichol, S. (2007) N-linked glycosylation of Gn (but not Gc) is important for Crimean Congo hemorrhagic fever virus glycoprotein localization and transport. *Virology*. **361**(2): 348-355
- Felitsyn, N., Kitova, E.N. and Klassen, J.S. (2001) Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behaviour. *Anal. Chem.* **73**: 4647-4661
- Ffrench-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T., Daborn, P.J., Bowen, D., and Blattner, F.R. (2000). A genomic sample sequence of the entomopathogenic bacterium *Photorhabdus luminescens* W14: Potential implications for virulence. *Appl. Environ. Microbiol.* **66**(8): 3310-3329.
- Ffrench-Constant, R.H., Waterfield, N., Daborn, P., Joyce, S., Bennett, H., Au, C., Dowling, A., Boundy, S., Reynolds, S., and Clarke, D. (2003). *Photorhabdus*: towards a functional genomic analysis of a symbiont and pathogen. *FEMS Microbiology Reviews* **26**(5): 433-456.
- Filmus, J. (2001) Glypicans in growth control and cancer. *Glycobiol.* **11**(3):19R-23R
- Finan, T.M., Kunkel, B., De Vos, G.F. and Signer, E.R. (1986). Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. *J. Bacteriol.* **167**: 66–72.
- Fischer, D., and Kissel, T. (2001) Histochemical characterization of primary capillary endothelial cells from porcine brains using monoclonal antibodies and fluorescein isothiocyanate-labelled lectins: implications for drug delivery. *Eur. J. Pharm. Biopharm.* **52**: 1 – 11.
- Fischer-Le Saux, M., Viallard, V., Brunel, B., Normand, P., and Boemare, N. E. (1999). Polyphasic classification of the genus *Photorhabdus* and proposal of new taxa: *P. luminescens* subsp. *luminescens* subsp. nov., *P. luminescens* subsp. *akhurstii* subsp. nov., *P. luminescens* subsp. *laumondii* subsp. nov., *P. temperata* sp. nov., *P. temperata* subsp. *temperata* subsp. nov. and *P. asymbiotica* sp. nov. *Int J Syst Bacteriol* **49**(4): 1645-1656.

- Forst, S., Dowds, B., Boemare, N., and Stackebrandt, E. (1997) *Xenorhabdus* and *Photorhabdus* spp. : bugs that kill bugs. *Annu. Rev. Microbiol.* **51**: 47-72
- Fuster, M.M., and Esko, J.D. (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat. Rev. Cancer.* **5**:562-542
- Gabius, H. (2001) Eukaryotic Glycosylation and lectins; Hardware of the sugar code (glycode) in biological information transfer. *Elect. J. Pathol. Histol.* **7**(1): 5-14
- Galili, U., Rachmilewitz, E.A., Peleg, A., and Flechner, I. (1984). A unique natural human IgG antibody with anti-alpha-galactosyl specificity. *J. Exp. Med.* **160**(5): 1519-1531
- Garber, N., Guempel, U., Belz, A., Gilboa-Garber, N., and Doyle, R.J. (1992) On the specificity of the D-galactose-binding lectin (PA-I) of *Pseudomonas aeruginosa* and its strong binding to hydrophobic derivatives of D-galactose and thiogalactose. *Biochim Biophys Acta.* **1116**: 331-333
- Gerrard, J.G, Joyce, S.A., Clarke, D.J., Ffrench-Constant, H., Nimmo, G.R., Looke, D.F., Feil, E.J., Pearce, L., and Waterfield, N.R. (2006) Nematode symbiont for *Photorhabdus asymbiotica*. *Emerging Infectious Diseases.* **12**(10): 1562-1564
- Gerrard, J.G., McNevin, S., Alfredson, A., Forgan-Smith, R., and Fraser, N. (2003) *Photorhabdus* species: bioluminescent bacteria as emerging human pathogens? *Emerging Infectious Diseases.* **9**(2): 251-254
- Geyer, H., and Geyer, R. (2006) Strategies for analysis of glycoprotein glycosylation. *Biochim. Biophys. Acta.* **1764**: 1853-1869
- Gilboa-Garber, N. (1972a). Purification and properties of hemagglutinin from *Pseudomonas aeruginosa* and its reaction with human blood cells. *Biochim. Biophys. Acta* .**273**(1): 165-73.
- Gilboa-Garber, N. (1972b). Inhibition of broad spectrum hemagglutinin from *Pseudomonas aeruginosa* by D-galactose and its derivatives. *FEBS Lett.* **20**(2): 242-244.
- Gilboa-Garber, N. (1982) *Pseudomonas aeruginosa* lectins. *Methods Enzymol.* **83**: 378-385

- Gilboa-Garber, N. and D. Sudakevitz (1999). The hemagglutinating activities of *Pseudomonas aeruginosa* lectins PA-IL and PA-IIL exhibit opposite temperature profiles due to different receptor types. *FEMS Immunol. Med. Microbiol.* **25**(4): 365-9.
- Gilboa-Garber, N., Katcoff, D., and Garber, N.C. (2000). Identification and characterization of *Pseudomonas aeruginosa* PA-IIL lectin gene and protein compared to PA-IL. *FEMS Immunol. Med. Microbiol.* **29**(1): 53.
- Gilboa-Garber, N., Mizrahi, L., and Garber, N. (1972). Purification of the galactose-binding hemagglutinin of *Pseudomonas aeruginosa* by affinity column chromatography using sepharose. *FEBS Lett.* **28**(1): 93-95.
- Glick, J., and Garber, N. (1983) The intracellular localization of *Pseudomonas aeruginosa* lectins. *J. Gen. Microbiol.* **129**: 3085-3090
- Gornik, O. and G. Lauc (2007). Enzyme linked lectin assay (ELLA) for direct analysis of transferrin sialylation in serum samples. *Clin. Biochem.* **40**: 718-723
- Goto, M. (2007) Protein O-glycosylation in fungi: Diverse structures and multiple functions. *Biosci. Biotechnol. Biochem.* **71**: 1415-1427
- Grass, S., Buscher, A.Z., Swords, W.E., Apicella, M.A., Barenkamp, S.J., Ozchlewski, N., and St. Geme, J.W., III (2003) The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis *Mol. Microbiol.* **48**: 737-751
- Gu, X., Xie, L., Harmon, B.J., and Wang, D.I. (1997) Influence of primatone RL supplementation on sialylation of recombinant human interferon- γ produced by Chinese hamster ovary cell culture using serum-free media. *Biotechnol. Bioeng.* **56**(4):353-360
- Gubareva, L.V., Kaiser, L., and Hayden, F.G. (2000). Influenza virus neuraminidase inhibitors. *Lancet* **355**(9206): 827-35.
- Hahn, H. (1997). The ype 4 pilus is the major virulence-associated adhesion of *Pseudomonas aeruginosa*. *Gene* **192**: 99-108.

- Hakomori, S. (2002) Glycosylation defining cancer malignancy: new wine in an old bottle. *Proc. Natl. Acad. Sci. USA*. **99**(16):10231-10233
- Haltner, E., Easson, J., and Lehr, C.M. (1997) Lectins and bacterial invasion factors for controlling endo- and transcytosis of bioadhesive drug carrier systems. *Eur. J. Pharm. Biopharm.* **44**: 3– 13.
- Han, R., and Ehlers, R. (2000). Pathogenicity, development, and reproduction of *Heterorhabditis bacteriophora* and *Steinernema carpocapsae* under axenic in vivo conditions. *J. Invertebr. Pathol* **75**: 55-58.
- Hang, H.C., and Bertozzi, C.R. (2005) The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg. Med. Chem.* **13**: 5021-5034
- Hazes, B., Boodhoo, A., Cockle, S.A., and Read, R.J. (1996) Crystal structure of the pertussis toxin-ATP complex: a molecular sensor. *J. Mol. Biol.* **258**: 661–671
- Hirschberg, C. B. (2001). Golgi nucleotide sugar transport and leukocyte adhesion deficiency II. *J. Clin. Invest.* **108**(1): 3-6.
- Hollister, J., Grabenhorst, E., Nimstz, M., Conradt, H., and Jarvis, J.L. (2002) Engineering the N-glycosylation pathway in insect cells for production of biantennary, complex N-Glycans. *Biochem.* **41**:15093-15104
- Holloway, C., and Cowen, J. (1997). Development of a scanning confocal laser microscopic technique to examine the structure and composition of marine snow. *Limnol. Oceanogr.* **42**:1340-1352
- Holmner, A., Askarieh, G., Okvist, M., and Krengel, U. (2007) Blood group antigen recognition by *Escherichia coli* heat-labile enterotoxin. *J. Mol. Biol.* **371**(3): 754-764
- Hooker, A.D., Goldman, M.H., Markham, N.H., James, D.C., Ison A.P., Bull, A.T., Strange, P.G., Salmon, I., Baines, A.J., and Jenkins, N. (1995) N-glycans of recombinant human interferon- γ change during batch culture of Chinese hamster ovary cells. *Biotechnol. Bioeng.* **48**(6):639-648
- Hooper, L. V. and J. I. Gordon (2001). Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity. *Glycobiol.* **11**(2): 1R-10R.

- Hsu, K., Gildersleeve, J.C., and Mahal, L.K. (2008) A simple strategy for the creation of a recombinant lectin microarray. *Mol. BioSyst.* **4**:654-662
- Hu, K., Li, J., and Webster, J.M. (1999) Nematicidal metabolites produced by *Photorhabdus luminescens* (Enterobacteriaceae), bacterial symbiont of entomopathogenic nematodes. *Nematology*. **1**: 457-469
- Huang, L., Hollingsworth, R.I., Castellani, R., and Zipser, P. (2004) Accumulation of high molecular weight amylose in Alzheimer's disease brains. *Glycobiol.* **14**(5):409-416
- Hölemann, A., and Seeberger, P.H. (2004) Carbohydrate diversity: synthesis of glycoconjugates and complex carbohydrates. *Curr. Opin. Biotechnol.* **15**:615-622
- Inbar, M. and L. Sachs (1969). Interaction of the carbohydrate-binding protein concanavalin A with normal and transformed cells. *Proc Natl Acad Sci U S A* **63**(4): 1418-25.
- Inoue, H., Nojima, H., and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*. **96**: 23-28
- Iskratsch, T., Braun, A., Paschinger, K., and Wilson, I.B. (2009) Specificity analysis of lectins using remodeled glycoproteins. *Anal. Biochem.* **386**(2):133-146
- Johansen, P.G., Marshall, R.D., and Neuberger, A. (1961) The preparation and some of the properties of a glycopeptides from hens egg albumin. *Biochem. J.* **78**:518-527
- Josenhans, C., Ferrero, R.L., Labigne, A. and Suerbaum, S. (1999). Cloning and allelic exchange mutagenesis of two flagellin genes of *Helicobacter felis*. *Mol Microbiol.* **33**: 350–362
- Joyce, S.A., Watson, R.J., and Clarke, D. (2006) The regulation of pathogenicity and mutualism in *Photorhabdus*. *Curr. Opin. Micirbiol.* **9**: 127-132
- Jurchen, J.C. and Williams, E.R. (2003) Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homo-dimers. *J. Am. Chem. Soc.* **125**:2817.

- Kamerling, J.P., Rijkse, I., Maas, A.M., van Kuik, A., and Vliegthart, J.F. (1988) Sulfated *N*-linked carbohydrate chains in porcine thyroglobulin. *FEBS Letters*. **241**(1,2):246-250
- Karaveg, K., Liu, Z., Tempel, W., Doyle, R., Rose, J., Wang, B. (2002). Crystallization and preliminary X-ray diffraction studies of lectin-1 from *Pseudomonas aeruginosa*. *Acta Crystallogr. D Biol. Crystallogr* **59**: 1241-1242
- Kasper, M., Migheli, A., Gold, L.R., and White, L.R. (1993) Embedding of lung tissue for immunoelectron microscopy. *Acta Histochem.* **95**: 221– 227.
- Kawaguchi, T., and Decho A.W. (2000). Biochemical characterization of cyanobacterial extracellular polymers (EPS) from modern marine stromatolites. *Preparative BioChemistry and BioTechnology* **30**: 321-330
- Kijne, J., Bauchrowitz, M., and Diaz, C. (1997) Root lectins and *Rhizobia*. *Plant Physiol.* **115**: 869-873
- Kim, H.J., Lee, S.J., and Kim, H. (2008) Antibody-based enzyme-linked lectin assay (ABELLA) for the sialylated recombinant human erythropoietin present in culture supernatant
- Kirkeby, S., and Moe, D. (2005). Analyses of *Pseudomonas aeruginosa* lectin binding to α -galactosylated glycans. *Current Microbiology* **50**: 309-313.
- 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 receptors in various mouse tissues. *Microbial Pathogenesis.* **40**(5): 191-197
- Kirkeby, S., Wimmerová, M., Moe, D., and Hansen, A.K. (2007). The mink as an animal model for *Pseudomonas aeruginosa* adhesion: binding of the bacterial lectins (PA-IL and PA-IIL) to neoglycoproteins and to sections of pancreas and lung tissues from healthy mink. *Microbes and Infection* **9**(5): 566.
- Klein, H., Löschner, B., Zyto, N., Pörtner, M., and Montag, T. (1998). Expression, purification and biochemical characterization of a recombinant lectin of *Sarcocystis muris* (Apicomplexa) cyst merozoites. *Glyconj J.* **15**: 147-153

- Kopp, K., Schlüter, M., and Werner, R.G. (1996) Monitoring the glycosylation pattern of recombinant interferon- ω with high pH anion exchange chromatography and capillary electrophoresis. *Drug Research*. **46**(12):1191-1196
- Kornfeld, R., and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides. *Ann. Rev. Biochem.* **54**: 631–664
- Kostlanova, N., Mitchell, E., Lortat-Jacob, H., Oscarson, S., Lahmann, M., Gilboa-Garber, N., Chambat, G., Wimmerova, M and Imberty, A. (2005) The fucose binding lectin from *Ralstonia solanacearum*: A new type of β -propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose and plant xyloglucan. *J. Biol Chem* **280**(30): 27839-27849
- Kovach, M., Elzer, P. (1995). Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**(1): 175-176.
- Laine, R.A. (1997) The information-storing potential of the sugar code; in Gabius H-J, Gabius, S. (eds): Glycosciences; Status and Perspectives. London-Weinheim Chapman & Hall, pp 1-14
- Lanne, B., Ciopraga, J., Bergstrom, J., Motas, C., and Karlsson, K. A. (1994) Binding of the galactose-specific *Pseudomonas aeruginosa* lectin, PA-I, to glycosphingolipids and other glycoconjugates. *Glycoconj. J.* **11**: 292-298
- Laughlin, R., Musch, M., Hollbrook, C., Rocha, F., Chang, E. and Alverdy, J. (2000). The key role of *Pseudomonas aeruginosa* PA-I lectin on experimental gut-derived sepsis. *Annals of Surgery* **232**(1): 133-142.
- Lavelle, E.C., Grant, G., Pusztai, A., Pfüller, U., and O'Hagan, D.T. (2000) Mucosal immunogenicity of plant lectins in mice. *Immunol.* **99**(1):30-37
- Lehr, C.M., and Pusztai, A. (1995) The potential of bioadhesive lectins for the delivery of peptide and protein drugs to the gastrointestinal tract, in: A. Pusztai, S. Bardocz (Eds.), *Lectins: Biomedical Perspectives*, Taylor and Francis, London, 1995, pp. 117–140.

- Lemaire, S., and Juillerat-Jeanneret, L. (2006) Glycosylation pathways as drug targets for cancer: glycosidase inhibitors. *Min. Rev. Med. Chem.* **6**(9):1043-1052
- Leriche, V., Sibille, P., and Carpentier, B (2000) Use of an enzyme-linked lectinsorbent assay to monitor the shift in polysaccharide composition in bacterial biofilms. *Appl. Environm. Biotechnol.* **66**(5):1851-1856
- Lin, J., Yao, J., Zhou, X., Sun, X., and Tang, K. (2003). Expression and purification of a novel mannose-binding lectin from *Pinellia ternate*, *Mol. Biotech.* **25**(3): 215-221
- Lis, H. and Sharon, N. (1998). Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chemical Reviews* **98**(2): 637-674.
- Liu, H, Pan, H.C., Cai, S.X., Chen, Z.W., Zheng, X.F., Yang, H.T., and Xiao, Z.Y. (2005) The effect of fermentation conditions on glycosylation of recombinant human interferon omega in yeast *Pichia pastoris*. *Chin. J. of Biotechn.* **21**(1):107-112
- Liu, Z.J., Tempel, W., Lin, D., Karaveg, K., Doyle, R.J., Rose, J.P., and Wang, B.C.. (2002) Structure determination of *P. aeruginosa* lectin-I using single wavelength anomalous scattering data from native crystals. *Am. Cryst. Assoc., Abstr. Papers* **29**: 98
- Lopez, M., Tetaert, D., Juliant, S., Gazon, M., Cerutti, M., Verbert, A., and Delannoy, P. (1999) *O*- Glycosylation of lepidopteran insect cell lines. *Biochim. Biophys. Acta.* **1427**(1): 49-61
- Marshall J.C., Christou N.V., and Meakins J.L. (1993) The gastrointestinal tract: the 'undrained abscess' of multiple organ failure. *Ann Surg.* **218**:111–119
- McCoy, J.P., Varani, J., and Goldstein, I.J. (1983). Enzyme-linked lectin assay (ELLA): use of alkaline phosphatase-conjugated Griffonia simplicifolia B4 isolectin for the detection of alpha-D-galactopyranosyl end groups. *Anal Biochem* **130**(2): 437-44.
- Mescher, M.F., and Strominger, J.L. (1976) Structural (shape-maintaining) role of the cell surface glycoprotein of *Halobacterium salinarium*. *Proc. Natl. Acad. Sci. U.S.A.* **73**(8):2687-3691

- Mescher, M.F., Strominger, J.L., and Watson, S.W. (1974) Protein and carbohydrate composition of the cell envelope of *Halobacterium salinarium*. *J. Bacteriol.* **120**: 945-954
- Messner, P. (1997). Bacterial glycoproteins. *Glycoconj. J.* **14**(1): 3-11.
- Mewe, M., Tielker, D., Schönberg, R., Schachner, M., Jaeger, K., and Schumacher, U. (2005). *Pseudomonas aeruginosa* lectins I and II and their interaction with human airway cilia. *J. Laryngol. Otol.* **119**(8): 595-9.
- Min, W., and Jones, D. (1992) Stability and detection of recombinant pre-pro-concanavalin A after cytoplasmic expression in *Escherichia coli*. *FEBS Lett.* **301**(3): 315-318
- Moorman, C., Benz, I., and Schmidt, A. (2002) Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesion heptosyltransferase of the AIDA system. *Infect. Immun.* **70**(5):2264-2270
- Mulvey, G., Kitov, P., Marcato, P., Bundle, D and Armstrong, G. (2001). Glycan mimicry as a basis for novel anti-infective drugs. *Biochimie* **83**: 841-847.
- Nakamura, T., Takada, N., Tono-zuka, T., Sakano, Y., Oguma, K., and Nishkawa, Y. (2007). Binding properties of *Clostridium botulinum* type C progenitor toxin to mucins. *Biochim et Biophysica Acta* **1770**: 551-555
- Neu, T. R., and Lawrence, J.R. (1999) Lectin-binding-analysis in biofilm systems. *Methods Enzymol.* **310**:145-152.
- Neu, T., Swerhone, T., and Lawrence, J. (2001). Assessment of lectin binding analysis for in-situ detection of glycoconjugates in biofilms systems. *Micobiol.* **147** (2): 299-313
- Opitz, L., Salaklang, J., Büttner, H., Reichl, U., and Wolff, M.W. (2007) Lectin-affinity chromatography for downstream processing of MDCK cell culture derived human influenza A viruses. *Vaccine.* **25**: 939-947
- Pilobello, K.T., Krishnamoorthy, L., Slawek, D., and Mahal, L.K. (2005) Development of a lectin microarray for the rapid analysis of glycoproteins. *Chem. Bio. Chem.* **6**:985-989

- Qiu, R., and Regneir, F.E. (2005) Use of multidimensional lectin affinity chromatography in differential glycoproteomics. *Anal. Chem.* **77**: 2802-2809
- Quandt, J., and Hynes, M. (1993). Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria. *Gene* **127**(1): 15-21.
- Rabinovich, G.A., Rubinstein, N., and Fainboim, L. (2002). Unlocking the secrets of galectins: a challenge at the frontier of glyco-immunology. *J. Leukoc. Biol.* **71**(5): 741-52.
- Richardson, W.H., Schmidt, T.M., and Nealson, K.H. (1988) Identification of an anthraquinone and a hydroxystilbene antibiotic from *Xenorhabdus luminescens*. *Appl Environ Microbiol.* **54**(6):1602-1605
- Rogerieux, F., Belaise, M., Terzidis-Travelsi, H., Greffard, A., Pilatte, A., and Lambrè, C.R. (1993). Determination of the sialic acid linkage specificity of sialidases using lectins in a solid phase assay. *Anal Biochem* **211**(2): 200-4.
- Ronin, C., Fenouillet, E., Hovsepian, S., Fayet, G., and Fournet, B. (1986) Regulation of thyroglobulin glycosylation. A comparative study of the thyroglobulins from porcine thyroid glands and follicles in serum-free culture. *J. Biol. Chem.* **261**: 7287-7293
- Rosenfeld, R., Bangio, H., Gerwig, G.J., Rosenberg, R., Aloni, R., Cohen, Y., Amor, Y., Plaschkes, I., Kamerling, J.P., and Maya, R.B. (2007) A lectin array method for the analysis of protein glycosylation. *J. Biochem. Biophys. Methods.* **70**: 415-426
- Ryan, M.P., Pembroke, J.T., and Adley, C.C. (2006) *Ralstonia pickettii*: a persistent Gram-negative nosocomial infectious organism. *J. Hospital Infection.* **62**:278-284
- Schaffer, C., and Messner, P. (2004) Surface layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. *Glycobiol.* **14**(8): 31R-42R
- Scharfman, A., Arora, S., Delmotte, P., Van Brussel, E., Mazurier, J., Ramphal, R., and Roussel, P. (2001). Recognition of Lewis x derivatives present on mucins by flagellar components of *Pseudomonas aeruginosa*. *Infection and Immunity* **69**: 5243-5248.

- Schuster, M., Lostroh, C., Ogi, T., and Greenberg, E. (2003) Identification, timing, and signal specificity of *Pseudomonas aeruginosa* quorum-controlled genes: a transcriptome analysis. *J Bacteriol.* **185**(7): 2066-2079
- Sharon, N. and H. Lis (2002). How Proteins Bind Carbohydrates: Lessons from Legume Lectins. *J. Agric. Food Chem.* **50**(22): 6586-6591.
- Sicard, M., Ferdy, J.B., Pagès, S., Le Brun, N., Godelle, B., Boemare, N., and Moulia, C. (2004) When mutualists are pathogens: an experimental study of the symbioses between *Steinernema* (entomopathogenic nematodes) and *Xenorhabdus* (bacteria). *J. Evol. Biol.* **17**: 985–993.
- Sicard, M., Hering, S., Schulte, R., Gaudriault, S., and Schulenberg, H. (2007). The effect of *Photorhabdus luminescens* (Enterobacteriaceae) on the survival, development, reproduction and behaviour of *Caenorhabditis elegans* (Nematoda: Rhabditidae). *Environ Microbiol* **9**(1): 12-25.
- Silva, C.P., Waterfield, N.R., Daborn, P.J., Dean, P., Chilver, T., Au, C.P., Potter, U., Reynolds, S.F., and Ffrench-Constant, R.H. (2002). Bacterial infection of a model insect: *Photorhabdus luminescens* and *Manduca sexta*. *Cell. Microbiol.* **4**(6): 329-339.
- Singh, R., Barden, A., Mori, T., and Beilin, L. (2001) Advanced glycation end-products: A review. *Diabetologia.* **44**(2):129-146
- Skehel, J. J. and D. C. Wiley (2000). Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin. *Annual Review Of Biochemistry* **69**: 531-569.
- Sleytr, U.B. and Thorne, K.J. (1976) Chemical characterization of the regularly arranged surface layers of *Clostridium thermosaccharolyticum* and *Clostridium thermohydrosulfuricum*. *J. Bacteriol.* **126**:869-882
- Smigielski, A., Akhurst, R., and Boemare, N. (1994) Phase variation in *Xenorhabdus nematophila* and *Photorhabdus luminescens*. *Appl Environ Microbiol.* **60**(1): 120-125
- Smith, A.C., de Wolff, J.F., Molyneux, K., Feehally, J., and Barratt, J. (2006) O-glycosylation of serum IgD in IgA nephropathy. *J. Am. Soc. Nephrol.* **17**:1192-1199

- Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. (1985) Measurement of protein concentration using bichinchonic acid. *Anal. Biochem.* **150**: 76-85
- Sonowane, W., Jyot, J., and Ramphal, R. (2006). *Pseudomonas aeruginosa* LecB is involved in pilus biogenesis and protease IV activity but not in adhesion to respiratory mucins. *Infection and Immunity* **74**(12): 7035-7039.
- Spector, S., Flynn, J.M, Tidor, B., Baker, T.A., and Sauer, R.T. (2003) Expression of N-formylated proteins in *Escherichia coli*. *Protein Expr. Purif.* **32**: 317-322.
- Spiro, R. G. (2002). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiol.* **12**(4): 43R-56.
- Spiro, R., and Bhoyroo, V. (1984) Occurrence of α -D-galactosyl residues in the thyroglobulins from several species. *J. Biol. Chem.* **259**(15):9858-9866
- Stancombe, P.R., Alexander, F.C.G., Ling, R., Matheson, M.A., Shone, C.C., and Chaddock, J.A. (2003) Isolation of the gene and large-scale expression and purification of recombinant *Erythrina cristagalli* lectin. *Prot. Express. Purific.* **30**(2):283-292
- Strauch, O., and Ehlers, R. (1998). Food signal production of *Photorhabdus luminescens* inducing the recovery of entomopathogenic nematodes *Heterorhabditis spp.* in liquid culture. *Appl. Microbiol. Biotechnol* **50**: 369-374.
- Sudakevitz, D., Kostlánová, N., Blatman-Jan, G., Mitchell, E.P., Lerrer, B., Wimmerová, M., Katcoff, D.J., Imberty, A., and Gilboa-Garber, N. (2004) *Mol. Microbiol.* **52**(3):891-700
- Sudakevitz, D., Levene, C., Sela, R., and Gilboa-Garber, N. (1996). Differentiation between human red cells of Pk and p blood types using *Pseudomonas aeruginosa* PA-I lectin. *Transfusion.* **36**(2):113-116
- Sumi, S., Arai, K., Kitihara, S., and Yoshida, K. (1999). Serial lectin affinity chromatography demonstrates altered asparagine-linked sugar-chain structures

- of prostate-specific antigen in human prostate carcinoma. *J. of Chrom. B: Biomed. Sci. and Appl.* **727**(1-2): 9.
- Sumner, J.B., and Howell, S.F. (1936) *J. Bacteriol.* **37**:227
- Szymanski, C.M., and Wren, B.W. (2005) Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Microbiol.* **3**: 225-237
- Szymanski, C.M., Burr, D.H., and Guerry, P.. (2002) *Campylobacter* protein glycosylation affects host cell interactions. *Infect. Immun.* **70**: 2242–2244
- Takata, S., Ohtani, O., and Watanabe, Y. (2000) Lectin binding patterns in rat nasal-associated lymphoid tissue (NALT) and the influence of various types of lectin on particle uptake in NALT. *Arch. Histol. Cytol.* **63**: 305– 312.
- Taylor, C. M. (1998). Glycopeptides and glycoproteins: Focus on the glycosidic linkage. *Tetrahedron* **54**(38): 11317-11362.
- Tchernychev, B., and Wilchek, M. (1996) Natural human antibodies to dietary lectins. *FEBS Lett.* **397**:139-142
- Thibault, P., Logan, S.M., Kelly, J.F., Brisson, J., Ewing, C.P., Trust, T.J., and Guerry, P. (2001) Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *J. Biol. Chem.* **276**(37): 34862-34870
- Thomas, G.M., and Poinar, G.O. (1979) *Xenorhabdus* gen. nov., a genus of entomopathogenic, nematophilic bacteria of the family *Enterobacteriaceae*. *Int. J. Syst. Bacteriol.* **29**: 352-360
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research.* **22**(22):4673-4680
- Tomana, M., Schrohenloher, R.E., Bennett, P.H., del Puente, A., and Koopman, W.J. (1994) Occurrence of deficient galactosylation of serum IgG prior to the onset of rheumatoid arthritis. *Rheumatol. Int.* **13**(6):217-220
- Turlin, E., Pascal, G., Roussell, J.C., Lenormand, P., Ngo, S., Danchin, A., and Derzelle, S. (2006) Proteome analysis of the phenotypic variation process in *Photorhabdus luminescens*. *Proteomics.* **6**: 2705-2725

- Van Damme, E.J., Balzarini, J., Smeets, K., Van Leuven, F., and Peumans, W. (1994). The monomeric and dimeric mannose-binding proteins from the *Orchidaceae* species *Listera ovata* and *Epipactis helleborine*: sequence homologies and differences in biological activities. *Glycoconj. J.* **11**(4): 321-32.
- Van den Brule, F., Califice, S., and Castronovo, V. (2004) Expression of galectins in cancer: A critical review. *Glycoconj. J.* **19**:537-542
- Van Weemen, B. K., and Schuurs A., H. (1971). Immunoassay using antigen-enzyme conjugates. *FEBS Lett.* **15**(3): 232-236.
- Virji, M. (1997) Post-translational modifications of type 4 pili and functional implications. *Gene* **192**: 141-147.
- Wada, Y., Azadi, P., Costello, C., Dell, A., Dwek, R., Geyer, H., Geyer, R., Kakehi, K., Karlsson, N., Kato, K., Kawasaki, N., Khoo, K., Kim, S., Kondo, A., Lattova, E., Mechref, Y., Miyoshi, E., Nakamura, K., Narimatsu, H., Novotny, M., Packer, N., Perreault, H., Katalinić, Pohlentz, Reinhold, V., Rudd, P., Suzuki, A., and Taniguchi, N. (2007) Comparison of the methods for profiling glycoprotein glycans-HUPO human disease glycomics/proteome initiative multi-institutional study. *Glycobiol.* **17**: 411-422
- Wang, P.H. (2005) Altered glycosylation in cancer: Sialic acids and sialyltransferases. *J. Cancer. Mol.* **1**: 73-81
- Ward, A., Sanderson, N.M., O'Reilly, J., Rutherford, N.G., Poolman, B., and Henderson, P. (2000) The amplification, expression, identification, purification, assay and properties of hexahistidine-tagged bacterial membrane transport proteins, p141-166. In Baldwin, S.A., (ed.) Membrane transport – a practical approach. Blackwell's Press: Oxford U.K.
- Watson, R.J., Joyce, S.A., Spencer, G.V., and Clarke, D. (2005) The *exbD* gene of *Photorhabdus temperate* is required for full virulence in insects and symbiosis with the nematode *Heterorhabditis*. *Mol. Microbiol.* **56**: 763-773
- Whitely, M., Lee, K.M., and Greenberg, E.P. (1999) Identification of genes controlled by quorum sensing in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA.* **96**: 13904-13909

- Williams, D.C., Lee, J.Y., Cai, M., Bewley, C.A., and Clore, G.M. (2005). Crystal Structures of the HIV-1 inhibitory cyanobacterial protein MVL free and bound to Man₃GlcNAc₂. *J. Biol. Chem.* **280**(32): 29269-29276
- Williams, J.C., Thomas, M., and Clarke, D. (2005b) The gene *stlA* encodes a phenylalanine ammonia-lyase that is involved in the production of a stilbene antibiotic in *Photorhabdus luminescens* TTO1. *Microbiol.* **151**: 2543-2550
- Winzer, K., Falconer, C., Garber, N.C., Diggle, S.P., Camara, M., and Williams, P. (2000). The *Pseudomonas aeruginosa* lectins PA-IL and PA-IIL are controlled by quorum sensing and by RpoS. *J. Bacteriol.* **182**(22): 6401-6411.
- Wu, A.M., Wu, J.H., Song, S., and Kabat, E.A (1996) *Bandeiraea (Griffonia) simplicifolia* lectin-I, isolectin A₄, reacting with Tn (GalNAcα1-3Ser/Thr) or galobiose (Galα1-4Gal) containing ligands. *FEBS Letts.* **398**: 183-186
- Wyss, C. (1998) Flagellins but not endoflagellar sheathproteins, of *Trepanema pallidum* and of pathogen-related oral spirochetes are glycosylated. *Infect. Immun.* **66**(12): 5751-5754
- Wyss, D., and Wagner, G. (1996) The structural role of sugars in glycoproteins. *Curr. Opin. Biotechnol.* **7**(4):409-417
- Xia, B., Royall, J.A., Damera, G., Sachdev, G.P., and Cummings, R.D. (2005) Altered O-glycosylation and sulfation of airway mucins associated with cystic fibrosis. *Glycobiol.* **15**(8):747-775
- Xu, J., and Hurlbert, R.E. (1990) Toxicity of irradiated media for *Xenorhabdus spp.* *Appl. Enviro. Microbiol.* **56**(3): 815-818
- Yabe., R., Suzuki, R., Kuno, A., Fukimoto, Z., Jigami, Y., and Hirabayashi, H. (2007) Tailoring a novel sialic acid-binding lectin from a ricin-B chain-like galactose-Binding protein by natural evolution-mimicry. *J. Biochem.* **141**:389-399
- Yang, Z.P. and Hancock W.S. (2004). Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J. Chrom. A.* **1053**(1-2): 79-88.

Zhou, X., Kaya, H., Heungens, K., and Goodrich-Blair, H. (2002) Response of ants to a deterrent factor produced by the symbiotic bacteria of entomopathogenic nematodes. *Appl. Environ. Microbiol.* **68**(12):6202-6209

Appendix I: Carbohydrate structures

Table 1: Symbol nomenclature for the representation of common monosaccharides.









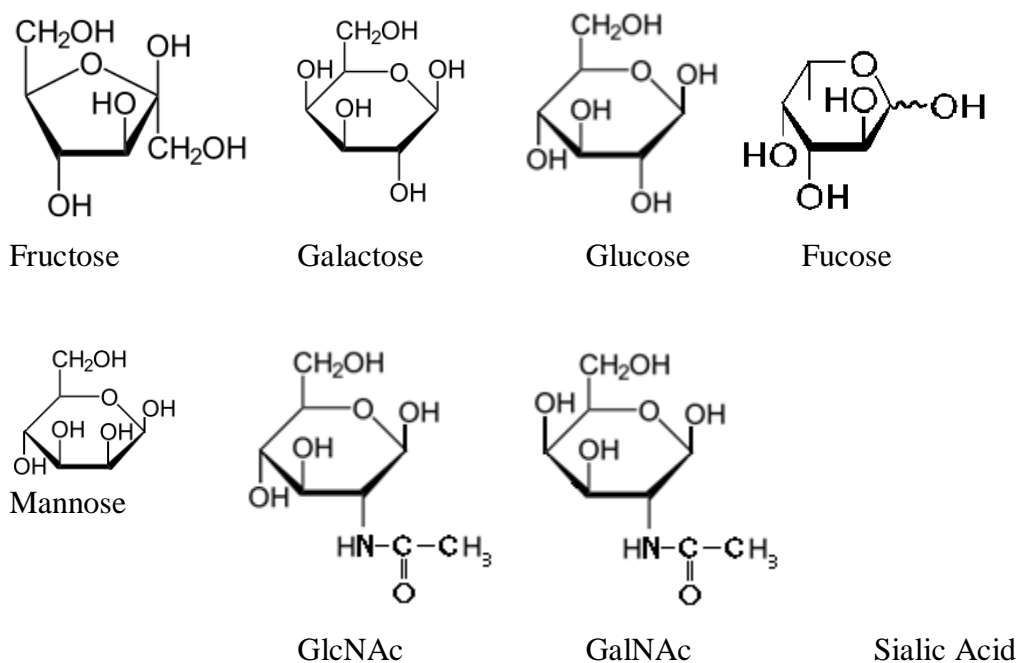
Symbol	Sugar	Abbreviation
	<i>N</i> -AcetylGlucosamine	GlcNAc
	Glucose	Glc
	<i>N</i> -AcetylGalactosamine	GalNAc
	Galactose	Gal
	Mannose	Man
	Fucose	Fuc
	Neuraminic Acid/Sialic Acid	Neu5Ac/Neu5Gc
	Fructose	Fru

Fig 1. Structures of some monosaccharides (Drawn using ChemBiodraw)



Phenyl- β -Galactose

Nitrophenyl- β -Galactose

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 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 own work.

Signed: _____

I.D. No: 51322729

Date: 02/09/2009

Acknowledgments

I would like to take this opportunity to offer my greatest thanks to my family.

Without your support, this thesis would not have been possible.

To my supervisors, Mick and Brendan, your patience, guidance and assistance are greatly appreciated. I have learned so much, I thank you both.

To my lab colleagues past and present, without your help this would not have been the enjoyable and treasured experience that it was. Hopefully, those dreaded words ‘I told you so’, will not be heard again.

Finally, my friends, thanks for the kind word and encouragement when things were tough.

Abbreviations

3D	Three dimensional
BCA	<i>Bicinchoninic acid</i>
bp	Base-pair
BSA	Bovine serum albumin
DBA	<i>Dolichos biflorus</i> agglutinin
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
ECL	<i>Erythrina cristagalli</i> lectin
EDTA	Ethylenediaminetetra acetic acid
FDA	The U.S. Food and Drug Administration
GSL-I	<i>Griffonia (Bandeiraea) simplicifolia</i> lectin I
IPTG	Isopropyl- β -D-thiogalactopyranoside
K _d	Dissociation constant
Log	Logarithm
LPS	Lipopolysaccharide
MAL-I	<i>Maackia amurensis</i> lectin I
MCS	Multiple cloning site
MW	Molecular weight,
OD	Optical density,
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein data bank
PNA	Peanut (<i>Arachis hypogaea</i>) agglutinin
SBA	Soy Bean (<i>Glycine max</i>) agglutinin
SNA	<i>Sambucus nigra</i> agglutinin-I
TEMED	N,N,N,N'-tetramethyl ethylenediamine
Tris	Tris (hydroxymethyl) amino methane
UEA-I	<i>Ulex europaeus</i> agglutinin-I
v/v	Volume per volume

w/v	Weight per volume
WGA	Wheat germ (<i>Triticum vulgaris</i>) agglutinin

Declaration	I
Acknowledgements	II
Abbreviations	III
Table of Contents	IV
List of Figures	X
List of Tables	
XV	
Abstract	
XVI	

Table of Contents

1.0 Introduction	1
1.1 Glycobiology.	2
1.2 Glycosylation.	2
1.2.1 <i>N</i> -linked glycosylation.	4
1.2.2 <i>O</i> -linked glycosylation.	7
1.2.3 Other glycoforms.	8
1.2.4 Bacterial glycosylation.	9
1.3 The importance of biological carbohydrates	11
1.3.1 Aberrant glycosylation in cancer cells.	12
1.4 Glycobiology and the biopharmaceutical industry.	13
1.5 Traditional glycoprotein analysis techniques	14
1.6 Lectins	16
1.7 Lectin families	17

1.7.1 Legume lectins	17
1.7.2 Cereal lectins	19
1.7.3 Amaryllidaceae and related family lectins	19
1.7.4 Galectins (formerly S-type lectins)	20
1.7.5 Complex carbohydrate binding molecules	21
1.7.6 Bacterial and viral lectins	22
1.8 Lectin applications	24
1.8.1 The enzyme linked lectin assay (ELLA)	25
1.8.2 Lectin arrays	26
1.8.3 Lectin affinity chromatography (LAC)	27
1.8.4 Lectin delivery molecules	28
1.8.5 Eukaryotic versus prokaryotic lectins	30
1.8.6 Lectin mutagenesis	31
1.9 The galactophilic lectin PA-IL	32
1.9.1 Discovery	32
1.9.2 Sequence analysis of the <i>lecA</i> gene	33
1.9.2 The structure of PA-IL	34
1.9.3 PA-IL affinity	37
1.9.4 PA-IL and pathogenicity	40
1.9.5 Other PA-IL associated functions	42
1.10 Project aims and objectives	43
2.0 Materials and Methods	44

2.1 Bacterial strains, primers and plasmids	45
2.2 Microbiological media	52
2.3 Solutions and buffers	53
2.4 Antibiotics	57
2.5 Storing and culturing of bacteria	57
2.6 Plasmid preparation by the 1-2-3 method	58
2.7 Plasmid preparation using the GenElute plasmid miniprep kit	58
2.8 Preparation of Gram-negative bacterial genomic DNA	59
2.9 Agarose gel electrophoresis	60
2.10 Isolation of DNA from agarose gels	61
2.11 Preparation of high efficiency competent cells	61
2.12 Transformation of competent cells	62
2.13 Determination of competent cell efficiency	
62	
2.14 Protein expression	63
2.14.1 Cell lysate preparation	63
2.14.2 Colony blot procedure	64
2.15: Immobilised metal affinity chromatography (IMAC)	65
2.15.1 IMAC using Ni-NTA resin	65
2.15.2 IMAC using FPLC and Amersham nickel-resin	65
2.15.3: Desalting of purified protein using HiPrep 26/10 desalting column	
66	
2.16 Recharging of IMAC resin	67
2.17 Protein quantification by BCA assay	67

2.18 SDS-PAGE.	67
2.18.1 Sample preparation	68
2.18.2 Sample application	68
2.18.3 Gel analysis	69
2.19 Native-PAGE	70
2.20 Western blot	71
2.21 Gel filtration	72
2.22 Blood preparation for hemagglutination assay	72
2.23 Hemagglutination assay	73
2.23.1 Hemagglutination inhibition assay	73
2.24 Enzyme linked lectin assay	73
2.25 Lectin purification using sepharose-4B	74
2.26 Immobilisation of protein onto cyanogen bromide-activated agarose	75
2.27 Tri-parental mating	76
2.28 In silico analysis of DNA and protein sequences	76
2.29 TA cloning of PCR products	77
2.30 Enzymatic reactions	78
2.31 ElectroSpray ionization mass spectrometry	79
2.32 <i>P. luminescens</i> pathogenicity assay	80
2.33 <i>P. luminescens</i> symbiosis assay	81

3.0 Cloning Expression and Purification of Recombinant PA-IL and PL-IL

83

3.1 Introduction	84
------------------	----

3.2 Cloning, small-scale expression and mutagenesis of lectin genes	84
3.2.1 Cloning and small scale expression of the <i>P. aeruginosa</i> gene <i>lecA</i> encoding PA-IL	84
3.2.2 Cloning and small-scale expression of <i>P. luminescens</i> gene <i>plu2096</i> which encodes PL-IL	87
3.2.3 Mutagenesis of PA-IL, and small-scale expression of the mutated gene encoding PA-IL _{mut1}	89
3.3 Expression of cloned lectin genes	92
3.3.1 Sub-cloning of lectin genes into pQE vectors	92
3.3.2 Selection of an <i>E. coli</i> expression strain for lectins	96
3.3.3 Optimization of expression conditions for lectins	101
3.4 Purification of recombinant lectins	105
3.4.1 Purification of untagged lectins	105
3.4.2 Purification of recombinant lectins by IMAC	109
3.4.3 Protein purification by FPLC	114
3.5 Total yields of recombinant lectin	117
3.6 Discussion	119

4.0 Structural Analysis of Recombinant PA-IL, PA-IL_{mut1} and PL-IL

124

4.1 Introduction	125
------------------	-----

4.2 Biochemical properties of native PA-IL and PL-IL	125
4.3 Construction of a standard curve for Superdex 75 column	127
4.4 Investigation of lectin compatibility with Superdex 75 GPC column	130
4.5 Determination of lectin size using Superdex 75 GPC	132
4.6 Validation on Superdex GF results using a different column matrix	136
4.7 Determination of PA-IL _{mut1} quaternary structure by GPC	143
4.8 Determination of lectin quaternary structure by mass spectrometry	144
4.9 Disruption of quaternary structure by pH and salt concentration	153
4.10 Determination of the valency of lectins by hemagglutination assay	156
4.11 Discussion	157

5.0 Determination of the Sugar Specificity of Recombinant Lectins

	161
5.1 Introduction	162
5.2 Determination of lectin specificity by hemagglutination inhibition	163
5.3 Establishment of the enzyme linked lectin assay (ELLA)	167
5.4 Quaternary structure affects lectin detection in ELLA	171

	5.5 Investigation into recombinant lectin specificity by ELLA	172
	5.6 Investigation of recombinant lectin affinity	184
190	5.7 Characterization of biopharmaceutical products by recombinant lectins	
	5.8 Lectin affinity chromatography	192
	5.9 Discussion	195
	 6.0 Determination of the role of PL-IL in the life-cycle of <i>P. luminescens</i>	
		199
	6.1 Introduction	200
	6.2 The bacterium <i>Photorhabdus luminescens</i>	200
	6.2.1 Mutualism in <i>P. luminescens</i>	202
	6.2.2 Glycan profile of <i>P. luminescens</i> associated organisms	204
	6.3 Construction of the plu2096::kan ^R mutant	205
	6.4 The role of PL-IL in <i>P. luminescens</i> pathogenicity	208
	6.5 The role of PL-IL in <i>P. luminescens</i> symbiosis	210
	6.5.1 Competitive symbiosis assay	211

6.6 Discussion	214
----------------	-----

7.0 Conclusions and Recommendations	217
--	-----

8.0 References	222
-----------------------------	-----

Appendix	243
----------	-----

List of Figures

Chapter 1

Fig 1.1 The precursor oligosaccharide unit for <i>N</i> -glycosylation.	5
Fig 1.2 Structure of the lipid molecule Dolichol	5
Fig 1.3 Biosynthesis of precursor oligosaccharide for <i>N</i> -linked glycosylation	6
Fig 1.4 Glycosylation of the mucin proteins in eukaryotic organisms	8
Fig 1.5 Basic core structure of GPI-anchors	9
Fig 1.6 Image of the glycan layer surrounding an endothelial cell	15
Fig 1.7 Monomeric structure of the lectin ConA	18
Fig 1.8 Crystal structure of the WGA lectin	19
Fig 1.9 Crystal structure of the GNA lectin	20
Fig 1.10 Crystal structure of the fungal galectin CGL-2	21
Fig 1.11 Source of available lectin 3D-crystal structures studied to date	24
Fig 1.12 Sequence alignment of PA-IL and PL-IL	34
Fig 1.13 Crystal structure of PA-IL complexed with galactose	35
Fig 1.14 Distribution of hydrophilic and hydrophobic regions on PA-IL	35
Fig 1.15 The A-D Interface of PA-IL	36
Fig 1.16 The C-termini of PA-IL	36
Fig 1.17 The location of cysteine residues on PA-IL	37
Fig 1.18 Glycan Binding Profile of PA-IL	38
Fig 1.19 3D Structure of the PA-IL sugar Binding Site	39

Chapter 2

Fig 2.1 pCR2.1 vector	48
Fig 2.2 pQE60 vector	48
Fig 2.3 pQE30 vector	49
Fig 2.4 pBBRMCS-5 vector	49
Fig 2.5 pJQ200sk+ vector	50
Fig 2.6 1 kb Molecular marker	60
Fig 2.7 NEB Prestained marker and Wide Range SigmaMarker	69

Fig 2.8 Schematic of western blot	70
Fig 2.9 Principle of TA cloning	77
 Chapter 3	
Fig 3.1 <i>P. aeruginosa</i> <i>lecA</i> coding sequence	85
Fig 3.2 Cloning strategy for pLecA1	86
Fig 3.3 Expression of PA-IL in <i>E. coli</i> JM109	87
Fig 3.4 Sequence analysis of the <i>P. luminescens</i> <i>plu2096</i> coding sequence	88
Fig 3.5 Expression of PL-IL in <i>E. coli</i> JM109	88
Fig 3.6 Schematic of Phusion™ site directed mutagenesis	90
Fig 3.7 Outline of the construction of PA-IL _{mut1}	90
Fig 3.8 Structure of PA-IL binding site and PA-IL _{mut1} binding site	91
Fig 3.9 Sequence alignment of PA-IL and PA-IL _{mut1}	91
Fig 3.10 Expression of PA-IL _{mut1} in JM109	92
Fig 3.11 Schematic of sub-cloning of <i>lecA</i> , <i>lecA_{mut1}</i> and <i>plu2096</i> into commercial pQE-expression vectors	95
Fig 3.12 Expression of PA-IL30 in BL21, KRX and XL10Gold	97
Fig 3.13 Expression of PL-IL in <i>E. coli</i> BL21, KRX and XL10Gold	98
Fig 3.14 Growth curves for the expression of recombinant PA-IL30 and PL-IL30 in <i>E. coli</i>	98
Fig 3.15 Effect of colony selection on lectin expression in <i>E. coli</i> BL21	99
Fig 3.16 Degradation of <i>E. coli</i> XL-10Gold expressed PA-IL _{mut1} 30 visualised by ES-MS	100
Fig 3.17 Effect of varying IPTG concentrations on expression of PL-IL in BL21	102
Fig 3.18 Effect of varying IPTG concentrations on growth rate of <i>E. coli</i> BL21	102
Fig 3.19 Expression of PA-IL in <i>E. coli</i> KRX	103
Fig 3.20 Expression of PL-IL30 in <i>E. coli</i> BL21	104
Fig 3.21 Purification of untagged PA-IL by affinity chromatography	

using sepharose-4B	105
Fig 3.22 Purification of untagged PA-IL using sepharose 4B	106
Fig 3.23A&B Attempted purification of PL-IL _{wt} and PA-IL _{mut130}	
by affinity chromatography using sepharose 4B	107
Fig 3.24 Effect of increasing temperature on <i>E. coli</i> cell lysate	
containing over-expressed untagged PL-IL _{wt}	108
Fig 3.25 Effect of acetic acid on purification by heating	108
Fig 3.26 Size fractionation of heat treated cell lysate	109
Fig 3.27 Purification of PA-IL30 with Qiagen Ni-NTA resin	110
Fig 3.28 Purification of PA-IL60 with Amersham Ni-NTA resin	111
Fig 3.29 Purification of PA-IL30 with Amersham Ni-NTA resin	111
Fig 3.30 Purification of PL-IL60 with Amersham Ni-NTA resin	112
Fig 3.31 Purification of PL-IL30 with Amersham Ni-NTA resin	112
Fig 3.32 Purification of PA-IL60 over HisTrap FF crude column by FPLC	115
Fig 3.33 Purification of PA-IL60 by FPLC-IMAC	115
Fig 3.34 Purification of PL-IL60 over HisTrap FF crude column by FPLC	116
Fig 3.35 Purification of PL-IL60 by FPLC-IMAC	116
Fig 3.36 Sequence alignment of entire PA-IL-like protein super-family	120

Chapter 4

Fig 4.1 Monomeric structure of PL-IL compared to PA-IL as predicted	
by the 3D-Jigsaw Protein Comparative Modelling Server	126
Fig 4.2 Development of size exclusion chromatography standard curve	128
Fig 4.3 Elution of protein standards on superdex 75 column	129
Fig 4.4 Schematic view of a superdex 75 bead	130
Fig 4.5 The adsorption A_{280nm} of PA-IL through the Superdex 75 10/300	
column in the presence/absence of raffinose	131
Fig 4.6 The adsorption A_{280nm} of PL-IL through the Superdex 75 10/300	
column in the presence of raffinose	131
Fig 4.7 Elution profile of PA-IL using the superdex 75 column	133
Fig 4.8 Elution profile of PL-IL in the superdex 75 column	134

Fig 4.9 Monomeric and polymeric structure of methyl methacrylate	137
Fig 4.10 Determination of column efficiency	137
Fig 4.11 Elution of protein standards from the Toyopearl HW-50S GPC column	138
Fig 4.12 Development of size exclusion chromatography standard curve for the Toyopearl HW50S GPC column	139
Fig 4.13 Elution profile of PA-IL in 0.2 M PBS pH 7.2	140
Fig 4.14 OD 280nm elution profile of PL-IL in 0.2M PBS pH 7.2	141
Fig 4.15 Comparison of PA-IL30 and PA-IL _{mutl} 30 elution profiles	143
Fig 4.16 Mass spectra of PA-ILwt investigated using a standard source	146
Fig 4.17 Dissociation of PA-IL tetramer ions into trimer, dimer and monomer ions by CID	147
Fig 4.18 Dissociation of PL-IL60 tetramer ions into trimer, dimer, and monomer ions by CID	148
Fig 4.19 Dissociation of the PL-IL30 tetramer ions into trimer and monomer ions by increasing collision voltage	149
Fig 4.20 Dissociation of PL-ILwt tetramer ions into trimer, dimer and monomer ions by CID	150
Fig 4.21 Elution volumes of PA-ILwt under varying salt and pH	154
Fig 4.22 Elution volumes of PA-IL60 under varying pH and salt	155
Fig 4.23 Illustration of hemagglutination by tetrameric and dimeric lectins	160

Chapter 5

Fig 5.1 Hemagglutination inhibition of PA-IL and PL-IL using simple glycans	165
Fig 5.2 Effect of temperature on hemagglutination activity	166
Fig 5.3 Comparison of blocking reagents in ELLA	169
Fig 5.4 Comparison of native and denatured glycoprotein immobilisation	169
Fig 5.5 Calcium is required for PL-IL and PA-IL _{mutl} binding	170
Fig 5.6 Binding of differently tagged forms of PA-IL and PL-IL to immobilised hyaluronidase	171

Fig 5.7A Identification of glycoprotein targets for PA-IL by ELLA	174
Fig 5.7B Identification of glycoprotein targets for PL-IL by ELLA	174
Fig 5.7C Identification of glycoprotein targets for PA-IL _{mut1} by ELLA	175
Fig 5.8 Investigation of lectin binding to asialylated glycoproteins	175
Fig 5.9 Binding of lectins to neuraminidase treated glycoproteins	176
Fig 5.10 Glycan structures that have been found on fetuin	178
Fig 5.11 Determination of recombinant lectin specificity by analysis of binding to glycosidase treated fetuin	179
Fig 5.12 Determination of recombinant lectin specificity by analysis of binding to glycosidase treated thyroglobulin	182
Fig 5.13 The structure of the most abundant <i>N</i> -linked glycan chain present in porcine thyroglobulin	183
Fig 5.14 Determination of lectin affinity for asialothyroglobulin	185
Fig 5.15 Determination of recombinant lectin affinity for asialofetuin	186
Fig 5.16 Inhibition of PA-IL binding to asialothyroglobulin by sugars	187
Fig 5.17 Inhibition of PL-IL binding to asialothyroglobulin by sugars	188
Fig 5.18 Inhibition of PA-IL _{mut1} binding to asialothyroglobulin by sugars	189
Fig 5.19 Binding of recombinant lectins to a biopharmaceutical product	191
Fig 5.20 Activation of sepharose by cyanogen bromide and protein coupling to the activated matrix	192
Fig 5.21 Location of surface lysine residues on the PA-IL molecule	193
Fig 5.22 Determining the amount of PA-IL _{mut1} which successfully immobilised on activated sepharose	193
Fig 5.23 Immobilisation of PA-IL _{mut1} to activated sepharose	194
Fig 5.24 Application of asialothyroglobulin onto PA-IL _{mut1} column	194

Chapter 6

Fig 6.1 The life cycle of <i>P. luminescens</i> and <i>H. bacteriophora</i>	202
Fig 6.2 The pUC4K vector	206
Fig 6.3 Schematic of homologous recombination events	207
Fig 6.4 Confirmation of the <i>plu2096::kan^R</i> mutant by PCR	208

Fig 6.5 Pathogenicity of <i>P. luminescens</i> strains towards <i>Galleria</i> larvae	209
Fig 6.6 Comparison of <i>P. luminescens</i> wild-type and the mutant <i>plu2096::kan^R</i> in a competitive pathogenicity assay	209
Fig 6.7 Infection of <i>G. mellonella</i> larvae by <i>P. luminescens</i>	210
Fig 6.8 Colonization of <i>Heterorhabditis</i> nematodes by wild-type and mutant strains	211
Fig 6.9 Proportions of strains on lipid agar plates in absence of nematodes	213
Fig 6.10 Proportion of each bacteria type present in the colonized nematodes in the competition assay at day 24	213

List of Tables

Chapter 1

Table 1.1 Summary of animal lectins	22
Table 1.2 Lectins studied for the purpose of drug targeting	29
Table 1.3 Summary of identified PA-IL targets	38

Chapter 2

Table 2.1 Bacterial strains	45
Table 2.2 Plasmids	46
Table 2.3 Primer Sequences	47
Table 2.4 SDS-PAGE gel recipes	68
Table 2.5 Silver Staining of SDS-PAGE gels	70

Chapter 3

Table 3.1: Summary of lectin nomenclature used in this study	94
Table 3.2 Protein yields obtained from optimised expression conditions	118

Chapter 4

Table 4.1 Construction of a protein molecular weight standard curve for the Superdex 75 high performance column	128
Table 4.2 Average molecular mass determined from Superdex 75 column	135
Table 4.3 Construction of a protein molecular weight standard curve for the Toyopearl HW50S GPC column	139
Table 4.4: Lectin sizes calculated using Toyopearl HW-50S column	142
Table 4.5: Summary of lectin quaternary structures observed by MS	151
Table 4.6 Comparison of ES-MS results with those observed by GPC	152
Table 4.7 Quantities of lectin required for hemagglutination	156

Chapter 5

Table 5.1 Structure of common oligosaccharides	164
Table 5.2 Hemagglutination inhibition against a range of sugars	165

Abstract

The growing development and application of glycoproteins as biopharmaceutical therapeutics has led to increased interest in the glycan content and glycoforms of glycoproteins. The level of glycosylation of these therapeutics have important implications for their efficacy. Lectins are naturally occurring bio-recognition protein molecules that bind glycoproteins. They have huge potential as specific tools in the analysis and purification of glycoproteins that are produced for therapeutic purposes. Lectins are found in all classes of organisms including the bacteria *Pseudomonas aeruginosa* and *Photorhabdus luminescens*. These bacteria respectively encode the highly studied lectin PA-IL and its previously uncharacterised homologue, PL-IL.

In this study these two lectins were expressed in *Escherichia coli* and purified to homogeneity via polyHis affinity tags which were fused to their *N*- and *C*- termini. It was shown through gel permeation chromatography and electrospray-ionisation Mass Spectrometry that the addition of affinity tags affected the quaternary structure of PA-IL to a greater extent than PL-IL. This was subsequently found to affect sugar binding activity. The sugar binding specificities for both lectins were determined by hemagglutination inhibition and enzyme linked lectin assays (ELLAs). Both lectins were found to have similar preferences for glycans terminating in α -Gal linkages. This specificity, in the case of PA-IL, could be altered through site directed mutagenesis, giving an insight into the roles of specific residues within the sugar binding pocket. Immobilisation studies on the recombinant lectins was carried out for their insertion onto novel analytical platforms for the ultimate characterisation and purification of therapeutic glycoproteins. The biological role of PL-IL within the organism *P. luminescens* was also investigated through pathogenicity and symbiosis assays.

Structural and Functional Characterisation of Lectins from the PA-IL Superfamily

Kenneth McMahon

Ph.D

2009

**Structural and Functional
Characterisation of
Lectins from the PA-IL Superfamily**

Thesis submitted for the degree of
Doctor of Philosophy

by
Kenneth McMahon, B.Sc.

Supervised by

Michael O'Connell B.Sc., Ph.D
and
Brendan O'Connor B.Sc., Ph.D

School of Biotechnology
Dublin City University
Ireland

September 2009