

1. INTRODUCTION

Lectins are carbohydrate-binding proteins of non-immune origin that reversibly bind sugar moieties of glycoconjugates without enzymatically modifying them (Lis and Sharon, 1998). Consequently, lectins are widely employed in biochemical research as carbohydrate-specific reagents (Adar *et al.*, 1998; Sharon and Lis, 2003).

All cell membranes and cell walls contain glycoconjugates whose constituents differ from one cell type to the other and can be cross-linked (agglutinated) by sugar-binding proteins. Lectins can therefore be used to probe cell surfaces and to fractionate cells, cell organelles and sub-cellular components (Yim *et al.*, 2001). Traditionally lectins have been used for blood typing as well as mitogenic stimulation of lymphocytes (Sharon and Lis, 1990; Kilpatrick, 1995). Moreover, since protein-carbohydrate interaction is the language of intercellular communication, lectins present excellent study models for biocommunication (Bouckaert *et al.*, 1999).

Although lectins are widely distributed in nature, plant lectins are the most abundant. Plant lectins are found mainly in the seeds in protein bodies and may constitute up to 10% of the total protein. The lectins are also found in the vegetative parts such as the stem, bark, roots and leaves, and more than one lectin may be present in the same tissue (Etzler, 1985).

Amongst the plant lectins, legume lectins are the largest and best-characterized group. Comparisons of both the amino acid sequences and the three-dimensional structures of the legume lectins show a very high degree of homology, yet they exhibit very distinct differences in their sugar-binding specificities (Sharon and Lis, 1990). For this reason, this large family of homologous proteins is highly suitable for use as a model system for studying the principles behind protein-carbohydrate recognition (Bouckaert *et al.*, 1999).

Legume lectins typically consist of dimers or tetramers of subunits with molecular masses 25 to 30 kDa that are held together by non-covalent interactions. Each subunit has one sugar-binding site and also binds calcium and manganese ions that are required for the sugar-

binding activity (Sharon and Lis, 1990). The sugar-binding site consists of four loops designated A, B, C and D. Conserved amino acid residues aspartate and glycine belonging to loops A and B respectively, as well as asparagine and an aromatic amino acid or leucine belonging to loop C, are required at the combining site. Loop D provides additional interactions, and in mannose/glucose-specific lectins its size is invariant and is possibly the determinant of the monosaccharide specificities in legume lectins (Sharma and Surolia, 1997; Hamelryck *et al.*, 1999; Loris *et al.*, 2004). It would therefore be interesting to establish if loop D is responsible for sugar-binding specificity, thereby providing a basis for systematically producing lectins with desired specificities.

A number of legume lectin genes have been cloned, sequenced and expressed in heterologous systems (Sharon, 1994). Sequencing confirmed the high homology even at the nucleotide level. Expression of the cDNAs encoding these lectins in *Escherichia coli* (*E. coli*) has led to the production of unglycosylated biologically active lectins demonstrating that although most legume lectin precursors are glycosylated, the carbohydrate moieties are not required for biological activity (Van Damme *et al.*, 1998b).

The majority of legume lectins have been isolated from seeds of crop plants while information about lectins from seeds of wild plants is scarce (Guzmán-Partida *et al.*, 2004). Some species of the Leguminosae family only grow in the tropical and sub-tropical climate zones. These tropical plants, particularly those of the wild flora, could contain lectins of unknown and potentially interesting carbohydrate specificities and/or stability properties that might find new applications in glycoconjugate analysis and biological research in general (S. Beeckmans, personal communication).

The seeds of the leguminous hardwood *Pterocarpus angolensis* (mukwa) contain a mannose/glucose-specific lectin. Interest in the mukwa seed lectin was generated from a study of haemagglutinins of trees and shrubs of Zimbabwe (Moore, 1979). This study showed that this lectin haemagglutinated only protease-treated human cord erythrocytes and not adult erythrocytes. Kaul and co-workers (1991) also observed that this seed lectin did not

agglutinate adult human erythrocytes even after trypsin treatment. These studies led to the conclusion that the lectin was specific for i-antigens as these are only expressed by fetal and neonatal red blood cells. The i-antigen decreases after birth as the I-antigen increases and the latter becomes fully expressed on adult red blood cells (Roelcke, 1995). Studies by Manyumwa (1998), however, showed that the mukwa seed lectin is not specific for cord erythrocytes but binds to the mannose core of immature cells.

Mukwa seeds are obtained seasonally (Van Wyk and Van Wyk, 1997) and the mukwa seed lectin has since been purified, characterized at the biochemical and molecular level (Manyumwa, 1998; Maramba, 1998) and its three-dimensional structure determined (Loris *et al.*, 2003; 2004). The seed lectin crystallizes readily and structural data for this lectin in complex with various carbohydrates is already available (Loris *et al.*, 2003; 2004). This work could be complemented by site-directed mutagenesis, which allows predictions about the structural and functional roles of particular amino acids in a protein to be rigorously tested in the laboratory. Expressing the cDNA encoding the lectin found in these seeds would make the recombinant lectin readily available to potential users as well as facilitating protein engineering of the mukwa seed lectin for comparative studies.

1.1 Aim of the Present Study

Protein-carbohydrate interactions are involved in diverse regulatory processes, hence understanding the principles that govern carbohydrate binding to proteins is essential. For such studies, legume lectins have been used as a model system because this family encompasses the widest variety of sugar specificities while maintaining a conserved tertiary structure. Site-directed mutagenesis of the carbohydrate-binding site of legume lectins has led to the identification of conserved amino acid residues required for sugar binding (Van Eijsden *et al.*, 1992; Adar and Sharon, 1996). However, the molecular basis of the differential sugar-binding characteristic of legume lectins is still poorly understood (Sharma *et al.*, 1998). While

the ligand-binding activity has been successfully altered (Yamamoto *et al.*, 2000b; Yim *et al.*, 2001), there is no rational approach to producing lectins with distinct and desired sugar specificities. Loop D, which is highly variable in length, conformation and sequence, yet completes the basic framework of monosaccharide recognition by a legume lectin, has been implicated as the major determinant of sugar specificity (Sharma and Surolia, 1997; Hamelryck *et al.*, 1998, Loris *et al.*, 2003). Consequently, loop D has been termed the variable loop or the specificity loop (Loris *et al.*, 1998).

The long-term goal of this study is to combine results from X-ray crystallography and site-directed mutagenesis to gain a better understanding of the determinants of sugar-binding specificity in legume lectins. The mannose/glucose specific lectin from mukwa seeds has been purified and extensively characterized (Loris *et al.*, 2003) and is a suitable model for this study. The closely related galactose-specific lectin from *Erythrina corallodendron*, ECorL (Adar *et al.*, 1989) and the 'promiscuous' *Ulex europaeus* II lectin, UEA II (Loris *et al.*, 2000) should allow for a comparative analysis of the structural basis for change in specificity. Like the mukwa tree, *Erythrina corallodendron* (coral tree) is deciduous and occurs throughout the tropics and sub-tropics. The dimeric ECorL is one of the most studied lectins (Stancombe *et al.*, 2003). The UEA II has the long version of the metal-binding loop just like the mukwa seed lectin yet the sugar specificities of these two lectins are unrelated (Loris *et al.*, 2004).

The project was based on predictions from computer modeling of the lectin crystal structures. When parts of loop D of ECorL and UEA II were separately superimposed onto loop D of the mukwa seed lectin the latter seemed to maintain its conformation (R. Loris, personal communication). In addition, certain substitutions and/or deletions of one or two amino acid residues of the mukwa seed lectin specificity loop were expected to change the specificity of this lectin from α to β -mannose/glucose (Loris *et al.*, 2004). The specific objectives were therefore to express the mukwa seed lectin in *E. coli* and to carry out *in vitro* mutagenesis in order to establish whether or not loop D is the sole determinant of sugar specificity in legume lectins.

2. LITERATURE REVIEW

2.1 Historical Background

The era of lectins started in 1888 when Herman Stillmark, at the Medical School of Dorpat in Estonia, observed that a castor bean (*Ricinus communis* L) plant extract named ‘ricin’ had haemagglutinating properties (Franz, 1988). Ever since then, there has been many significant historical landmarks in the studies of lectins.

A decade after Stillmark’s discovery, the proteins/glycoproteins that are now called lectins were given the name ‘haemagglutinins’ because of their ability to agglutinate erythrocytes (Elfstrand, 1898, cited by Van Damme *et al.*, 1998a). These proteins, however, were observed to be capable of agglutinating other cells such as spermatozoa, bacteria, lymphocytes and yeast cells. Subsequently lectins were referred to as ‘agglutinins’ (Goldstein and Hayes, 1978).

In 1908, K. Landsteiner and H. Raubitschek demonstrated different haemagglutinating properties by various seed extracts, namely *Phaseolus vulgaris* (red kidney bean), *Pisum sativum* (pea), *Lens culinaris* (lentil) and *Vicia sativa* (vetch) (Sharon, 1977; Van Damme *et al.*, 1998a). In 1919 J.B. Sumner obtained a pure protein for the first time when he successfully isolated crystalline concanavalin A, a lectin from jack bean (*Canavalia ensiformis*), for which he was awarded a Nobel prize. Seventeen years later, Sumner and S.F. Howell reported that concanavalin A binds sugar moieties on cells and also precipitates glycogen from solution (Sharon and Lis, 1998).

In the 1940s, W.C. Boyd, R.M. Reguera and K.O. Renkonen demonstrated that some plant lectins selectively agglutinated particular erythrocytes within the human ABO blood system and were thus blood group-specific (Boyd and Ruguera, 1949).

The next two major developments were in the early 1960s. Firstly, P.C. Nowell (1960) found out that the lectin from the red kidney bean (*Phaseolus vulgaris*), traditionally known as phytohaemagglutinin, was mitogenic to resting lymphocytes. Secondly, J.C. Aub and co-

workers (1963) observed that the wheat germ agglutinin preferentially agglutinated malignant cells (Sharon and Lis, 1998).

The early 1970s saw the establishment of the three-dimensional structure of concanavalin A using high resolution X-ray crystallography by G.M. Edelman and co-workers and independently by K. Hardman and C.F. Ainsworth. In 1974, G. Ashwell and A.G. Morell identified the first mammalian lectin, hepatic asialoglycoprotein receptor specific for terminal galactose in serum glycoproteins. Two years later, Y. Reisner observed that the peanut agglutinin discriminates cortical from medullary cells in mice and together with Sharon used lectins for the fractionation of lymphocytes. Reisner and co-workers used the soybean agglutinin for purging of bone marrow for transplantation in 1981 (Sharon and Lis, 1998).

In 1988, K. Drickamer identified carbohydrate recognition domains in animal lectins, while in plant lectins, particularly in legume lectins, the sugar-lectin interaction has been extensively studied by various scientists from the 1990s to this present day.

2.2 Defining a Lectin

The term lectin, derived from the Latin verb 'legere', means to pick out or to choose (Sharon and Lis, 1998). Boyd first proposed this term to describe the selectivity/specificity of these proteins for some human blood group antigens (Boyd and Shyleigh, 1954).

Over the years, the definition of a lectin has changed, from being based on the functional activities of the protein to being based on its structure. Goldstein and co-workers (1980) defined a lectin as 'a carbohydrate-binding protein (or glycoprotein) of non-immune origin that agglutinates cells and/or precipitates glycoconjugates.' This definition distinguished this group of proteins from antibodies directed against carbohydrate moieties but encompassed sugar-specific enzymes with multiple binding sites. The definition was considered too restrictive as it confined lectins to only those proteins with at least two sugar-binding sites thereby excluding monovalent lectins such as mannose-binding proteins from

orchids (Van Damme *et al.*, 1994) and poor agglutinating toxins such as abrin (Kocourek and Horejši, 1983).

Kocourek and Horejši (1983) defined a lectin as ‘a protein of non-immune origin, capable of specific recognition and reversible binding to sugar moieties of complex carbohydrates without altering the covalent structure of any of the recognized glycosyl ligand.’ According to this definition, lectins can display various biological activities, as they can be toxins or hormones. However, carbohydrate-binding proteins that only bind free sugars such as some transport proteins, chemotaxis receptors and receptors of operons of enzymes involved in carbohydrate metabolism are not lectins.

After observations that some lectins contain a second type of binding site that interacts with non-carbohydrate ligands, Barondes in 1988 re-defined a lectin as ‘a carbohydrate-binding protein other than an enzyme or an antibody.’

Advances in molecular cloning of lectin and lectin-related genes advocated for an update of the definition of lectins. Firstly some plant enzymes such as class I plant chitinases and some type 2 ribosome inactivating proteins (RIPs) are fusion proteins consisting of a carbohydrate-binding domain tandemly arrayed with a separate catalytic domain. Secondly, evolutionarily and structurally related lectin-like proteins exist. These lectin-related proteins, such as the α -amylase inhibitor and arcelins found in several legume species, are devoid of carbohydrate-binding activity (Mirkov *et al.*, 1994). In view of these factors, Peumans and Van Damme (1995) defined a lectin as ‘a protein that possesses at least one non-catalytic domain that binds reversibly to a specific mono- or oligosaccharide.’ Their definition encompasses a broader range of proteins.

2.3 Lectins in Nature

Lectins are a general class of sugar-binding proteins and are most abundant in plants (Rüdiger, and Rougè, 1998). Plant lectins act as mediators of cellular recognition in a variety of systems such as the attachment of nitrogen-fixing bacteria to roots of leguminous plants (Díaz *et al.*, 1989).

In animals, lectins occur in small amounts and seem to be tissue-specific. Animal lectins appear to function in the clearance of glycoproteins from the circulatory system and in the intracellular translocation and targeting of glycoproteins (Sharon and Lis, 1989; Weis and Drickamer, 1996; Kilpatrick, 2002).

In microorganisms, lectins are located on the microbial cell surfaces. Microbial lectins mediate the sugar-specific adherence of microbes to host cells, which is a prerequisite for infection (Sharon and Lis, 1989).

2.4 Plant Lectins

Plant lectins mark the beginning of lectinology and have been extensively studied (Etzler, 1985; Lis and Sharon, 1986). This group of proteins is heterogenous as the lectins differ from each other with respect to their biochemical properties, molecular structure and sugar-binding specificities (Etzler, 1985).

Lectin-containing plants are found in many botanical groups including mono- and dicotyledons, molds and lichens. However, lectins have mainly been detected in Leguminosae and Gramineae (Etzler, 1986; Lis and Sharon, 1986). Plant lectins have been shown to occur in virtually all tissues and organs but their concentrations vary depending on the taxonomic grouping, species, developmental stage and specific location within the plant. Often large amounts of the lectin occur in storage organs such as seeds and bulbs but may also be present in the vegetative tissues such as the leaves of the same plant. Seed lectins are mainly localized in the cotyledonary parenchyma cells where they appear during the later stages of seed

maturation (Etzler, 1985). In some *Phaseolus* species, seed lectins constitute up to 50% of the total protein (Van Damme *et al.*, 1998b) yet barely detectable levels are present in leaves of leek (Van Damme *et al.*, 1993).

Many plant species contain more than one lectin in relatively different proportions. These lectins, which may occur in different tissues or at different developmental stages of the same plant, may be related, or may differ with respect to physico-chemical and sugar-binding properties (Lis and Sharon, 1986). For instance, in *Robinia pseudoacacia*, *Maackia amurensis* and *Sophora japonica*, the bark and seed lectins are different and are encoded by different genes (Van Damme *et al.*, 1998a). Multiple molecular forms of lectins (isolectins) with different sugar specificities have been isolated from *Cratylia mollis* seeds (Santos *et al.*, 2004). Isolectins, which can be encoded by the same gene or different genes, may exhibit differences in charge and amino acid sequence (Hoedemaeker *et al.*, 1994). The charge heterogeneity is due to the frequently imprecise C-terminal post-translational proteolysis (Van Driessche *et al.*, 1988; Young *et al.*, 1995).

Most plant lectins are glycoproteins as they contain covalently bound carbohydrates. The glycoproteins are of two types, those containing primarily mannose and N-acetylglucosamine such as the soybean agglutinin and those containing L-arabinose and galactose such as the potato lectin (Lis and Sharon, 1986; Young *et al.*, 1995). A number of lectins of the first type, for example the lectin from *Erythrina cristagalli* (Iglesias *et al.*, 1982), also contain L-fucose and xylose. The carbohydrate moieties are, however, not required for biological activity as chemically deglycosylated potato and tomato lectins retain their haemagglutinating activity and carbohydrate specificity (Lis and Sharon, 1986).

2.4.1 Classification

Initially, plant lectins were classified into specificity groups depending on the sugar that best inhibited either haemagglutination or polysaccharide/glycoprotein precipitation by the lectin (Goldstein and Poretz, 1986). Van Damme and co-workers (1998a) proposed a further

sub-division based on the overall structure of the lectin subunits. Advances in the analysis of the structure of plant lectins and their genes have now enabled lectins to be sub-divided into families of structurally and evolutionarily related proteins (Van Damme *et al.*, 1998a).

2.4.1.1 Based on Sugar Specificity

Customarily, the sugar specificity of lectins is examined by the Landsteiner hapten-inhibition technique. In this technique, different sugars are tested for their ability to inhibit the precipitin reaction between the lectin and a reactive macromolecule or the haemagglutination reaction (Sharon and Lis, 1990). Based on this criterion, six specificity groups have been identified (see Table 2.1). These groups are: mannose-, D-mannose/D-glucose-, D-galactose/N-acetyl-D-galactosamine-, N-acetyl-D-glucosamine (chitobiose)-, L-fucose-, and N-acetyl-neuramic (sialic) acid-binding (Van Damme *et al.*, 1998b).

Members of each group may differ with respect to anomeric specificity and affinity for various analogs or derivatives of the sugar. Some lectins such as concanavalin A and the pea lectin exhibit pronounced anomeric specificity whereas other lectins such as those from soybean and castor bean are indifferent (Goldstein and Poretz, 1986; Sharon and Lis, 1990). Many lectins tolerate variations at the C-2 position of the sugar to which they bind. Consequently, D-mannose-specific lectins usually also react with D-glucose and to some extent with N-acetyl-D-glucosamine. However, the configuration of the hydroxyl group at C-4 is critical as galactose-specific lectins do not react with glucose and vice versa (Sharon and Lis, 1990).

Some lectins have a higher affinity for specific di- and trisaccharides, and still others interact more strongly with the oligosaccharide than with the given sugar monomer. Generally, such lectins broadly recognize the oligosaccharides but the specific sugar monomer is usually at the non-reducing end (Sharon and Lis, 1990; Weis and Drickamer, 1996). Certain lectins react exclusively with oligosaccharides while others bind glycopeptides better than the corresponding oligosaccharides. These lectins that bind exclusively to the complex (modified)

oligosaccharide side chains of typical animal glycoproteins belong to the ‘complex specificity’ group (Sharon and Lis, 1990; Loris *et al.*, 1998) (see Table 2.1).

Table 2.1: Examples of Plant Lectins Exhibiting Different Specificities

Specificity	Plant Name	Common Name	Reference
Mannose	<i>Heliathus tuberosus</i>	Jerusalem artichoke	Peumans <i>et al.</i> , 2000
	<i>Galanthus nivalis</i>	Snowdrop	Peumans <i>et al.</i> , 2000
D-mannose/D-glucose	<i>Lens culinaris</i>	Lentil	Foriers <i>et al.</i> , 1981
	<i>Pterocarpus angolensis</i>	Mukwa tree	Loris <i>et al.</i> , 2003
D-galactose/ N-acetyl-D-galactosamine	<i>Glycine max</i>	Soybean	Sharon and Lis, 1990
	<i>Erythrina corallodendron</i>	Coral tree	Shaanan <i>et al.</i> , 1991
N-acetyl-D-glucosamine (chitobiose)	<i>Solanum tuberosum</i>	Potato	Ayouba <i>et al.</i> , 1991
	<i>Triticum vulgare</i>	Wheat	Ayouba <i>et al.</i> , 1991
L-fucose	<i>Ulex europeus</i> I	Furze (gorse)	Thomas and Surolia, 2000
	<i>Lotus tetragonolobus</i>	Asparagus pea	Thomas and Surolia, 2000
N-acetyl-neuramic (sialic) acid	<i>Sambucus nigra</i>	Elderberry	Shibuya <i>et al.</i> , 1987
	<i>Maackia amurensis</i>	Maackia	Peumans and Van Damme, 1995
Complex	<i>Phaseolus vulgaris</i>	Kidney bean	Sharon and Lis, 1990
	<i>Griffonia simplicifolia</i> IV	(not known)	Sharon and Lis, 1990

2.4.1.2 Based on Overall Structure of Mature Lectin

Lectins are sub-divided into four structural groups, namely merolectins, hololectins, chimerolectins and superlectins (see Figure 2.1).

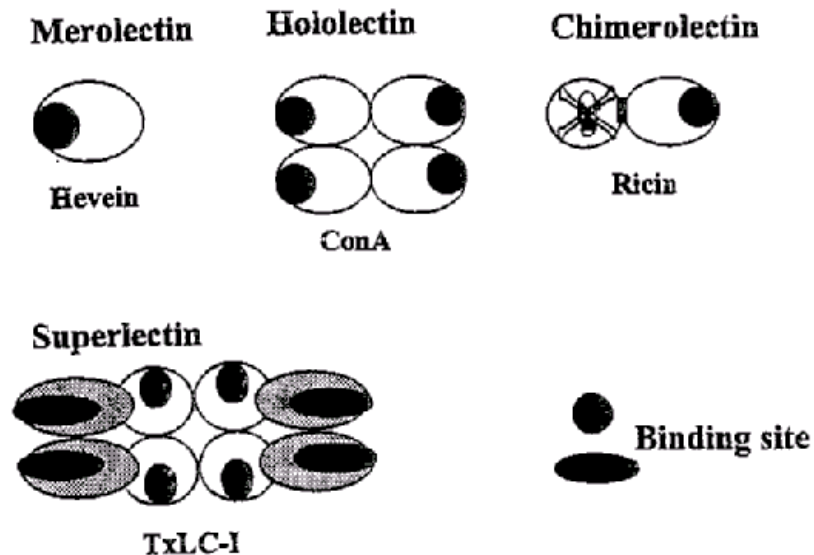


Figure 2.1: Schematic representation of the four structural types of plant lectins: merolectins, hololectins, chimerolectins and superlectins (from Van Damme *et al.*, 1998b). Con A is concanavalin A and TxLC-1 is the tulip lectin.

Merolectins are small proteins consisting of a single carbohydrate-binding domain (CBD) (Van Damme *et al.*, 1998b; Neumann *et al.*, 2004). These monovalent lectins are not able to agglutinate cells or precipitate glycoconjugates. Typical merolectins include the chitin-binding protein hevein from the latex of the rubber tree (*Hevea brasiliensis*) and the monomeric mannose-binding proteins from orchids (Van Damme *et al.*, 1998b).

Most plant lectins belong to the sub-group of hololectins. Hololectins contain at least two carbohydrate-binding domains that are either identical or highly homologous and consequently bind either the same or structurally similar sugars (Van Damme *et al.*, 1998b; Neumann *et al.*, 2004). Given they are di- or multivalent, hololectins are capable of agglutinating cells and/ or precipitating glycoconjugates (Van Damme *et al.*, 1998b).

Chimerolectins are fusion proteins composed of a carbohydrate-binding domain tandemly arrayed with an unrelated domain (Van Damme *et al.*, 1998b; Neumann *et al.*, 2004). The latter domain may have another biological activity, such as enzymatic activity, but acts independently of the carbohydrate-binding domain. Chimerolectins with multiple carbohydrate-binding sites, such as the type 2 ribosome inactivating proteins (RIPs), behave as hololectins. On the other hand, chimerolectins with one carbohydrate-binding site, such as the class I plant chitinases, behave as merolectins (Van Damme *et al.*, 1998b).

Superlectins are fusion proteins composed of two tandemly arrayed structurally and functionally different carbohydrate-binding domains (Van Damme *et al.*, 1998b; Neumann *et al.*, 2004). This special type of chimerolectin recognizes structurally unrelated sugars. The chimeric nature of superlectins is illustrated by the protomer of the tulip lectin TxLC-I that consists of an N-terminal mannose-binding domain tandemly arrayed with an unrelated N-acetyl-D-galactosamine-binding domain (Cammue *et al.*, 1986; Van Damme *et al.*, 1998b).

2.4.1.3 Based on Evolutionary Relationships

Analysis of available gene and protein sequences of lectins distinguishes four large and three small families of evolutionarily-related proteins. The major groups are the legume lectins (Sharon and Lis, 1990), the chitin-binding lectins (Raikhel *et al.*, 1993), the monocot mannose-binding lectins (Van Damme *et al.*, 1995b) and the type 2 ribosome inactivating proteins (RIPs) (Barbieri *et al.*, 1993). The small groups are the jacalin-related lectins, the amaranthin lectin family and the Cucurbitaceae phloem lectins (Van Damme *et al.*, 1998a).

Legume lectins are a large family of homologous proteins that are confined to the species of the plant family Leguminosae (Sharon and Lis, 1990). This family is of particular interest in the present study (see Section 2.5).

The chitin-binding lectins are composed of one or more hevein domains (Raikhel *et al.*, 1993). Hevein, which consists of a single polypeptide chain of forty-three amino acids, is rich in glycine residues and contains eight cysteine residues that are involved in interchain

disulphide bridges (Van Damme *et al.*, 1998a; Neumann *et al.*, 2004). Chitin-binding lectins have been isolated from several taxonomically unrelated plant families. Examples include the *Oryza sativa*, the *Solanum tuberosum* and the *Viscus album* lectins from the Gramineae, Solanaceae and Viscaceae families respectively (Van Damme *et al.*, 1998a). However, it must be noted that there are some chitin-binding legume lectins and Cucurbitaceae phloem lectins without hevein domains.

Monocot mannose-binding lectins are a superfamily occurring in six different families, namely Amaryllidaceae, Alliaceae, Araceae, Orchidaceae, Iridaceae and Liliaceae. These lectins occur exclusively in monocotyledonous species and have a marked sequence homology. Examples are the *Allium ursinum* and the *Allium sativum* (garlic) lectins (Van Damme *et al.*, 1998a; Zhao *et al.*, 2003).

Ribosome inactivating proteins (RIPs) catalytically inactivate eukaryotic ribosomes. Type 2 RIPs possess a highly specific rRNA N-glycosidase activity and exhibit carbohydrate-binding activity (Barbieri *et al.*, 1993; Van Damme *et al.*, 1998a). These lectins are a superfamily of structurally unrelated proteins that are found in members of taxonomically unrelated plant families. Examples include ricin, the *Abrus precatorius* and the *Sambucus nigra* lectins of the Euphorbiaceae, Fabaceae and the Sambucaceae families respectively (Van Damme *et al.*, 1998a).

Jacalin-related lectins comprise of all carbohydrate-binding proteins that share sequence similarity with the T-antigen specific agglutinin from jack fruit (*Artocarpus integrifolia*). This group is made up of a few galactose-specific lectins from the family Moraceae, namely *Artocarpus* and *Maclura*, as well as many mannose-binding lectins (Peumans *et al.*, 2000; Rougé *et al.*, 2003). The galactose-specific lectins, such as jacalin from jack fruit, are located in storage protein vacuoles and are tetramers composed of long α chains that are non-covalently associated to short β subunits. On the other hand, mannose-specific lectins such as the one found in rhizomes of hedge bindweed (*Calystegia sepium*) consist of

uncleaved protomers and are located in the cytoplasm (Van Damme *et al.*, 1998a; Peumans *et al.*, 2000; Rougé *et al.*, 2003).

Many Cucurbitaceae species such as the *Momordica charantia* and the *Trichosanthes cucumerina* are cultivated in many countries (Kenoth *et al.*, 2003). These Cucurbitaceae species have lectins in their phloem exudates. The lectin genes, for instance in pumpkin, are expressed in companion cells of phloem tissue (Bostwick *et al.*, 1992).

The Amaranthin group comprises of closely related N-acetyl-D-galactosamine-specific lectins isolated from various *Amaranthus* species where they are typically found in seeds. These lectins are homodimeric and the *Amaranthin caudatus* lectin is an example of lectins belonging to this group (Van Damme *et al.*, 1998a).

2.4.2 Physiological Functions

The widespread occurrence of plant lectins and their conservation during evolution suggests that they may have some fundamental biological role in the plant, yet their physiological function is still not clear (Etzler, 1986; Putszai, 1991). Specific carbohydrate binding properties of lectins have led to the proposal that they serve as recognition molecules within the cell, between cells and between organisms (Chrispeels and Raikhel, 1991; Peumans and Van Damme, 1993).

The bulk of lectins are localized in protein bodies in storage tissues of seeds, rhizomes, tubers, bulbs and bark. As a consequent of their location in cell vacuoles and their general pattern of synthesis, lectins are thought to be mainly storage proteins (Law, 2000). This proposal is supported by the fact that lectins are utilized by the germinating seed and are synthesized during seed development (Rüdiger, 1997). Rhizomes and bulbs also mobilize their reserves upon emergence of a new plant to provide the growing shoot with amino acids (Peumans and Van Damme, 1993).

Lectins may also play a role in packaging storage proteins in protein bodies (Rüdiger and Rougé, 1998). *In vitro* studies indicate that the pea seed lectin binds to the seed storage

protein vicilin and phospholipid monolayers of the membrane of protein bodies. This observation suggests that this lectin may serve as a link between other storage proteins and the inner surface of the protein storage vacuole in developing seed storage tissue (Higgins *et al.*, 1983b; Brewin and Kardailsky, 1997).

Lectins have also been implicated to have a role in the defense mechanisms of plants (Chrispeels and Raikhel, 1991; Peumans and Van Damme, 1993; Zhao *et al.*, 2003). Glycoconjugates such as chitin in the fungal cell wall, or fetuin and mucin found on the surface of the intestinal tract of plant predators such as nematodes, insects and higher animals are more complex than those common in plants. The fact that plant lectins are capable of recognizing and binding these glycoconjugates when seeds or other plant organs are attacked by microorganisms or eaten by plant predators supports their possible role in plant defence (Peumans and Van Damme, 1995; Duranti and Gius, 1997; Moore *et al.*, 2000).

Gatehouse and co-workers (1990) proposed that some lectins have insecticidal activity based on the high toxicity of ricin to the coleopterans *Callosobruchus maculatus* and *Anthonomus grandis*. When tested in artificial diets, the snowdrop (*Galanthus nivalis*) agglutinin has been identified to be toxic towards *homoptera* insects such as the rice brown planthopper (Zhao *et al.*, 2003). Moreover, the expression of the snowdrop lectin in transgenic tobacco plants resulted in the appearance of protection against aphids (Hilder *et al.*, 1995).

The chitin-binding lectins from the latex of the rubber tree, *Hevea brasiliensis* and from the seeds of the amaranth, *Amaranthus caudatus* have been definitely shown to inhibit fungal growth (Peumans and Van Damme, 1995; Rüdiger and Rougé, 1998). A lectin from *Talisia esculenta* seeds also interfere with fungal growth (Freire *et al.*, 2002). In addition, the anti-nutritional reactions associated with legume lectins reduce the digestibility of plant storage tissues by herbivorous predators and other plant pathogens (Sharon, 1994; Moore *et al.*, 2000). For instance, the common bean is not attractive to possible predators because the accumulation of large quantities of the *Phaseolus vulgaris* lectin, phytohaemagglutinin, leads to severe effects on the intestine of higher animals (Putzai, 1991).

Lectins are also thought to serve as mediators of the symbiosis between the nitrogen-fixing microorganisms, *Rhizobia*, and leguminous plants (Díaz *et al.*, 1989; Brewin and Kardailsky, 1997; Moore *et al.*, 2000). The experimental findings reported so far cannot be generalized yet these demonstrations strongly support this hypothesis. Firstly, the lectin in *Pisum* (pea) roots is responsible for specific recognition by the bacteria as demonstrated by the fact that *Trifolium* (clover) roots transformed with a *Pisum* lectin gene are nodulated by the *Pisum*-specific *Rhizobium* strain which normally does not nodulate clover (Díaz *et al.*, 1989; Sharon, 1994). Secondly, the peanut root agglutinin (PRA-II) from *Arachys hypogaea* was found on the surface of seven-day old peanut seedling roots, suggesting its possible role in the recognition of *Rhizobium* by the peanut root (Kalsi *et al.*, 1995). Thirdly, the accumulation of the lectin found in peanut nodule tissues, is dependent on factors related to nitrogen supply and demand (Law, 2000). Chrispeels and Raikhel (1991), however, speculate that the role of lectins in symbiosis may have evolved from their ability to agglutinate and immobilize bacteria as a defense mechanism.

2.4.3 Anti-Nutritional Properties

Plant-based foodstuffs such as salads, fruits, spices, dry cereals and roasted nuts contain appreciable but varying amounts of lectins. These lectins are responsible for the poor intolerance of many raw legumes, the tomato and wheat germ lectins by both humans and livestock (Liener, 1986; Duranti and Gius, 1997). Intolerance may be due to the appetite-depressing effects and/or resistance to gut proteolysis (Pusztai, 1991; Vasconcelos and Oliveira, 2004). Interaction of the lectin with the gut's microbial flora can result in a change in the composition of microorganisms or selective microbial overgrowth. The different toxic metabolites produced may eventually enter the systemic circulation resulting in anti-nutritional and mild allergic effects (Liener, 1986; Van Damme *et al.*, 1998b).

Pusztai and co-workers conducted a series of comprehensive studies (Liener, 1986; Pusztai *et al.*, 1993a) that definitely identified the lectin as the main factor responsible for the

toxicity associated with the consumption of *P. vulgaris* in its raw form. Animals fed on the purified phytohaemagglutinin showed that when ingested, this lectin is highly resistant to gut proteases. In addition, the lectin binds to brush border cells of the intestine and is rapidly endocytosed into the cells where it induces enhanced metabolic activity that eventually leads to overgrowth of the small intestine. Rats fed on diets containing genetically modified potatoes expressing the *Galanthus nivalis* lectin showed proliferation of the gastric mucosa (Ewen and Pusztai, 1999).

Normal cooking inactivates some lectins but others, such as the wheat germ lectin, survive heat treatment (Liener, 1986; Grant *et al.*, 1995). An outbreak of gastroenteritis has been reported in many instances following the consumption of partially cooked beans (Liener, 1986; Vasconcelos and Oliveira, 2004).

Morphological changes in the intestinal mucosa have been reported to accompany the ingestion of lectins present in a number of plants. These multivalent lectins interact with glycoconjugates on the surface epithelium (Pusztai, 1991) and this interaction appears to reduce digestion and impair absorption of other nutrients. Moreover, gut protein losses, most likely mucin, increase, leading to malnutrition and subsequent weight loss (Liener, 1986; Moore *et al.*, 2000).

Oral administration of low doses of some lectins may, however, have beneficial effects on the gut's immune system, absorptive efficiency, bacterial flora and endocrine system (Pusztai and Bardocz, 1997). For instance, the snowdrop lectin prevents colonization of the gut by *Escherichia coli* as it blocks the fimbrial intestinal receptors (Pusztai *et al.*, 1993b; Van Driessche *et al.*, 2000). Mistletoe (*Viscum album*) lectins, when administered to healthy adults, initiate an immunomodulating response as evidenced by a strong proliferation of peripheral blood mononuclear cells that is accompanied by increased production of both tumour necrosis factor α and interleukin-6 (Pryme *et al.*, 2004).

2.4.4 Purification and Detection

Most plant lectins are isolated from seeds as these are a cheap, rich and readily available source of lectins. The seeds are first de-coated as the seed coat usually contains tannins and other phenolic compounds that are likely to interfere with subsequent purification steps (Rüdiger, 1988). De-coated seeds are then ground to a fine seed meal. Lipids and coloured materials are removed from the seed meal using organic solvents such as hexane, methanol or petroleum ether (Goldstein and Poretz, 1986).

Generally, purification of the lectin begins with a saline or buffer extraction of the finely ground fat-free seed meal. Cell debris is removed by filtration through cheesecloth followed by centrifugation. Ammonium sulphate fractionation of the crude extract yields a lectin-containing precipitate that can be re-solubilized. Low pH and heating can be employed to remove acid-soluble and heat-stable lectins respectively. Most legume lectins require cations such as calcium and manganese in the extraction buffers if the proteins are to remain fully active (Rüdiger, 1988).

Although conventional techniques such as ion exchange and molecular exclusion chromatography have successfully been used to purify lectins, partially purified lectins are usually subjected to affinity chromatography on immobilized sugar or glycoconjugates. This technique exploits the specific sugar-binding capacity of the lectin, after which the adsorbed lectin is then specifically eluted by adding free ligand or acid (Rüdiger, 1988; Kennedy *et al.*, 1995).

Detection of lectins is normally by agglutination of fresh animal or human erythrocytes. The lectin, which usually has at least two sugar-binding sites, is able to cross-link cells by binding to the carbohydrate antigens expressed on the surface of red blood cells. This interaction leads to clumping of the latter and the cells form aggregates that, if large enough, are visible to the naked eye (Boorman *et al.*, 1988; Swain *et al.*, 1996). Activity of the lectin is expressed as titre, that is, the reciprocal of the maximal dilution of the lectin that gives visible

aggregation. It must be noted, however, that scoring of haemagglutination activity is subjective (Swain *et al.*, 1996).

In some cases, the erythrocytes are first digested with enzymes such as papain, neuraminidase or trypsin, to render them more sensitive to agglutination. Such treatment exposes carbohydrate structures that are normally embedded within the cell surfaces (Goldstein and Poretz, 1986; Machuka *et al.*, 1999).

Latex bead agglutination can be used to detect lectins, particularly in laboratories that lack facilities to handle blood products. In this technique, polystyrene latex beads coated with glycoconjugates such as ovalbumin and horseradish peroxidase substitute the erythrocytes, and bovine serum albumin, a non-glycosylated protein, is used as a control (Kaul *et al.*, 1991).

2.4.5 Applications

The specific and reversible carbohydrate-binding properties of lectins, coupled with the ease of their preparation, make these proteins invaluable tools in diverse areas of biological, biomedical and biochemical research.

Lectins immobilized on solid matrices act as ligands for affinity chromatography in the purification of sugar-containing macromolecules, making it possible to isolate glycoconjugates from complex biological mixtures (Kobata and Endo, 1992; Sharon, 1994). The high specificity of lectins allows the resolution of closely related compounds such as glycoforms of a glycoprotein or glycopeptides that differ only slightly in the structure of their carbohydrate chains (Sharon, 1994; Rao *et al.*, 1998).

Routinely, lectin-binding is employed to demonstrate that membrane receptors for hormones, growth factors, neurotransmitters, immunoglobulins and toxins are glycoconjugates (Sharon, 1994). Cells, cell organelles and sub-cellular components can be fractionated using lectins. For instance, mouse splenocytes can be fractionated into B and T cells using the soybean agglutinin that agglutinates the B subpopulation. The soybean agglutinin is also used

to remove immunocompetent lymphocytes from human bone marrow used for transplantation (Kilpatrick, 1995; Sharon and Lis, 1998; Yim *et al.*, 2001).

Traditionally, lectins are used as blood cell typing agents that identify and distinguish between markers on erythrocyte surfaces in the ABO blood grouping system. The lectin from *Lotus tetragonolobus* serves to identify blood type O cells while that from *Dolichos biflorus* binds more strongly to type A₁ than to the A₂ subgroup. The lectin from *Vicia graminea* is specific for blood type N and is used to differentiate between M and N cells (Sharon, 1994).

Lectins with known carbohydrate specificities can be used to characterize and localize oligosaccharides on cells. Cell surfaces can therefore be probed, using these lectins, to detect changes in carbohydrate patterns that occur during cell development and diseases such as cancer (Sumar, 1994). Consequently, lectins tagged with fluorescent dyes, enzymes, gold particles or biotin (Sharon and Lis, 1998), have gained wide application in both histochemistry and diagnostic histopathology (Gabius and Gabius, 1991; Brooks *et al.*, 1997; Ewen, 1998). For instance, since the demonstration that the peanut agglutinin binds malignant cells in preference to normal cells in breast glands, this lectin has been used widely as a probe to detect malignant cells in several tissues (Kilpatrick, 1995; Chacko and Appukuttan, 2001).

Binding of some lectins to cells, results in the activation of mitogenesis in lymphocyte populations (Sharon, 1977; Seshagirirao and Prasad, 1995; Singh *et al.*, 2004). Mitogenic lectins, such as concanavalin A and phytohaemagglutinin, have been used to characterize polyclonal activation of T-lymphocytes (Kilpatrick, 1995). Activated lymphocytes proliferate and secrete a broad range of cytokines in large amounts (Moore *et al.*, 2000). Mitogenic stimulation by lectins provides a simple way of assessing the immunocompetence of patients suffering from diseases like AIDS and monitoring the effects of immunosuppressive manipulations such as stress. In addition, since chromosomes are easily visualized in stimulated cells, mitogenic stimulation is also employed for the preparation of chromosome maps for karyotyping, sex determination and detection of chromosomal defects (Sharon, 1994).

The use of lectins as therapeutic agents, whereby lectins are used to target and deliver drugs to their site of action, is still at experimental stages. The rationale behind lectin-mediated drug targeting is that diseased cells such as transformed and cancerous cells often express different glycans compared with their normal counterparts. Consequently, lectins could be used as carriers to target drugs specifically to different cells and tissues (Beis *et al.*, 2004). In 1988, Wooley and Naisbett proposed the use of the tomato lectin to target the luminal surface of the small intestine whereby the lectin would demonstrate bioadhesion. Besides the targeting to specific cells, the lectin-sugar interaction can also be used to trigger vesicular transport into or across epithelial cells (Beis *et al.*, 2004).

Genes of some plant lectins have been expressed in transgenic plants in a bid to confer resistance to attack by insects and nematodes by the agricultural crop. For instance, the snowdrop lectin has been successfully expressed in transgenic tobacco (Hilder *et al.*, 1995) and rice (Foissac *et al.*, 2000) resulting in added protection of the plants against *Myzus persicae*, a peach-potato aphid and *Nephotettix virwscens*, a green leafhopper respectively. It is, however, advisable to produce transgenic plants with a combination of lectin genes and other toxic proteins such as chitinases, proteases and amylase inhibitors as predators tend to become tolerant to a single protectant (Sharon and Lis, 1998).

2.4.6 Lectin Genes

Many plants contain multiple lectin genes as evidenced by the fact that some plants have more than one lectin (Sharon and Lis, 1990; Etzler, 1992). These lectins can be expressed in the same tissue, or differentially expressed in different tissues and/or developmental stages (Jordan and Goldstein, 1994). For instance, the peanut (*Arachis hypogaea* L) was shown to originate from a multigene family in which the gene for the nodule lectin is different from that of the seed lectin (Law, 2000). The isolectins found in the dimeric wheat germ agglutinin are due to the random combination of different subunits, each produced by a separate genome in the polyploidy wheat (Etzler, 1992).

DNA sequence determination of the lectin genes has revealed the absence of introns (Van Damme *et al.*, 1998a; Srinivas *et al.*, 2000; Stancombe *et al.*, 2003). For instance, Hoffmann and Donaldson (1985), Yamauchi and Minamikawa (1990) and Sharma and co-workers (1996) showed that the respective genes encoding phytohemagglutinin from *Phaseolus vulgaris*, concanavalin A from *Canavalia gladiata* and the winged bean basic agglutinin (WBAI) from *Psophocarpus tetragonolobus* have no intervening sequences. Lack of introns explains why the lectin genes are functional even when they are expressed in bacteria.

Although lectins are found in numerous plant species of different taxonomic groupings, amino acid sequencing of the proteins as well as molecular cloning of the lectin cDNAs has been carried out mainly in legumes (Sharon and Lis, 1990; Van Damme *et al.*, 1995b). Complementary DNA clones have been used as probes to measure levels of lectin mRNA in legume tissues. Transcripts hybridizing with the cDNAs increase during the mid-maturation stage of development of seeds and decrease during the later stages (Higgins *et al.*, 1983a; Etzler, 1986; Machuka *et al.*, 1999). The increases in lectin mRNA indicate maximal production of lectins at this stage (Yamauchi and Minamikawa, 1990), suggesting that the accumulation of lectins in legume seeds may be regulated at the transcriptional stage (Etzler, 1985; Lis and Sharon, 1986).

Both the amino acid and nucleotide sequences of legume lectins show a high degree of homology (Sharon and Lis, 1990). To date, however, lectin genes from species belonging to many taxonomic groups have been cloned (Van Damme *et al.*, 1998b).

2.5 Legume Lectins

Legume lectins are a large family of homologous carbohydrate-binding proteins that are found exclusively in the plant family Leguminosae (Sharon and Lis, 1990; Loris *et al.*, 1998). However, not all lectins found in legumes are legume lectins. For instance, the class I

chitinase from *Phaseolus vulgaris* belongs to the chitin-binding lectins (Van Damme *et al.*, 1998a).

Generally legume lectins are classified into two groups, the one-chain and the two-chain lectins. The one-chain lectins are either dimers or tetramers of one type of subunit. The two-chain lectins are made up of two different subunits, the α - or light chain and the β - or heavy chain, held together by non-covalent forces (Van Driessche, 1988). Examples of the one-chain lectins are the soybean agglutinin and phytohaemagglutinin and those of the two-chain lectins are from pea and lentil seeds (Loris *et al.*, 1998).

Abundant quantities of legume lectins are found mainly in mature seeds with smaller amounts present in vegetative tissues such as leaves, stems, bark, roots and nodules. These lectins, which are broadly distributed and easily isolated, exhibit a wide variety of carbohydrate specificities in spite of their strong sequence and structural conservation (Sharon and Lis, 1990; Rini, 1995; Rego *et al.*, 2002). Consequently, legume lectins are the model system of choice for studying the molecular basis of recognition between proteins and carbohydrates (Moreno *et al.*, 1997; Loris *et al.*, 1998; Manoj *et al.*, 2000).

2.5.1 Biosynthetic Pathway

Legume lectins are synthesized on the endoplasmic reticulum (ER) as pre-proteins and subsequently enter the secretory pathway. The prolectins are converted into mature lectins by post-translational modifications (Higgins *et al.*, 1983a; Van Driessche, 1988).

A leader or signal sequence of about twenty amino acid residues precedes the chain of the mature lectin and guides its transport into the lumen of the ER. Co-translational processing involves the removal of the signal peptide by a signal peptidase and if glycosylation sites are accessible, oligosaccharides are N-linked by a glycosyltransferase (Lord, 1985; Van Driessche, 1988; Rüdiger, 1997; Van Damme *et al.*, 1998a). The prolectin is further processed during or after intracellular transport from the ER via the Golgi apparatus to the storage vacuoles. This

processing involves proteolytic cleavage of N-terminal, C-terminal or internal peptides of the polypeptide backbone as well as a series of complex glycan modifications arising from both the addition and subsequent removal of sugar residues (Lord, 1985; Van Driessche, 1988; Roug   *et al.*, 1991; Van Damme *et al.*, 1998b).

Analysis of legume lectin genes and their products indicate that both one-chain and two-chain lectins are derived from similar precursors (see Figure 2.2). The mRNA encoding one-chain lectins is converted into the prolectin by the co-translational removal of the signal peptide. The prolectin is then transported via the Golgi complex to the vacuoles where it is converted into the mature polypeptide by the removal of a C-terminal propeptide. Synthesis of the two-chain lectins takes place in a similar way except that the latter undergo an additional proteolytic processing on the carboxyl side of an asparagine residue by a specific vacuolar endoprotease (Lord, 1985; Van Driessche, 1988; Moore *et al.*, 2000). This processing gives two smaller polypeptides and in some cases, as in the pea lectin, is accompanied by the removal of a short internal peptide (Lord, 1985). The position of the cleavage site determines whether the resulting polypeptides are similar in size as in the *Cladrastis lutea* lectin, or dissimilar as in the *Vicia faba* lectin (Van Driessche, 1988; Van Damme *et al.*, 1998b).

Legume lectins belonging to the Diocleae tribe, genera *Canavalia* and *Dioclea*, however, undergo a unique, complex mode of protein processing from that described above (Van Driessche, 1988). This biosynthetic pathway was first described for *Canavalia ensiformis* (Con A) (Carrington *et al.*, 1985). This lectin is first synthesized as an inactive glycosylated pre-proprotein. Co-translational modification involves the removal of the signal peptide to give an inactive glycosylated prolectin, proCon A. An internal glycosylated peptide is then excised from proCon A and the resulting peptides are transposed and re-ligated. Finally, a tetrapeptide is removed from the N-terminus of the ligation product to give active Con A (Carrington *et al.*, 1985; Van Driessche, 1988; Roug   *et al.*, 1991; Van Damme *et al.*, 1998a).

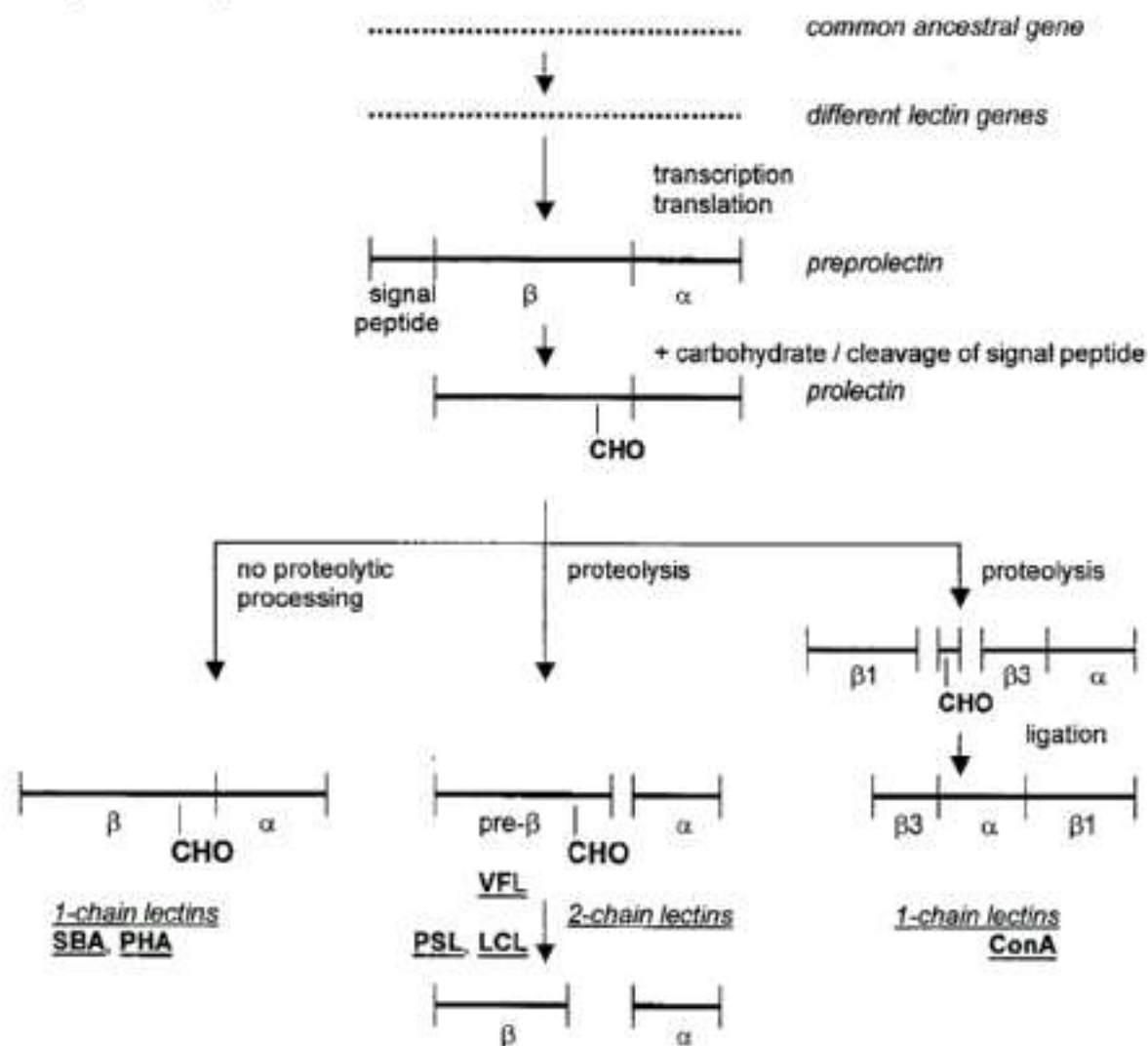


Figure 2.2: Schematic representation of the biosynthesis, co- and post-translational modifications of legume lectins (taken from Van Driessche, 1988). CHO stands for carbohydrate moiety, SBA for soybean agglutinin, PHA for phytohaemagglutinin from *Phaseolus vulgaris*, VFL for *Vicia faba* lectin, PSL for *Pisum sativum* lectin, LCL for *Lens culinaris* lectin, Con A for concanavalin A.

2.5.2 Structure

Typically, legume lectins are built up of two or four identical or highly homologous subunits that are held together by non-covalent interactions (Van Driessche, 1988; Van Damme *et al.*, 1998b). Consequently, legume lectins are sub-divided into dimeric and tetrameric forms according to the number of protomers (Rougè *et al.*, 1991; Van Damme *et al.*, 1998a). Each subunit has a molecular mass of 25 to 30 kDa and usually consists of a single

polypeptide chain of about 250 amino acid residues. In some instances, the polypeptide chain is partly cleaved into two smaller polypeptides that are either similar or dissimilar in size (Van Driessche, 1988; Van Damme *et al.*, 1998b).

Most legume lectins are glycosylated, containing a small number of N-linked carbohydrate units. These glycan chains can be the high mannose type or of the complex type. The soybean lectin contains an oligomannose side chain, whereas the *Phaseolus vulgaris* haemagglutinin contains both high-mannose and complex type glycans. The carbohydrate is, however, not required for the sugar-binding activities of the glycoprotein lectins (Sharon and Lis, 1990; Van Damme *et al.*, 1998b).

All legume lectins contain the divalent cations Ca^{2+} and a transition metal, usually Mn^{2+} , at specific metal-binding sites. The metal ions, which are found in each subunit, are held in place by interactions with specific amino acid residues and are essential for the carbohydrate-binding activity of the lectins (Sharon and Lis, 1990; Van Damme *et al.*, 1998b; Neumann *et al.*, 2004). These ions stabilize an unusual *cis*-peptide bond between an alanine and an aspartate residue such that the aspartate side chain is correctly positioned for sugar-binding (Hamelryck *et al.*, 1998; Loris *et al.*, 2004).

A number of legume lectins have been sequenced by chemical means or by deducing the amino acid sequences from the nucleotide sequence of cDNAs or genomic DNA fragments. Comparisons of the different amino acid sequences show that they all share a high degree of sequence similarity over the entire length of the lectin monomer and contain highly conserved regions (Strosberg *et al.*, 1986; Van Driessche, 1988; Sharon and Lis, 1990; Van Damme *et al.*, 1998a).

The legume monomer is a jelly-roll motif (see Figure 2.3) that comprises of a flat six-stranded back β -sheet and a curved seven-stranded front β -sheet joined by a small five-stranded β -sheet (S-sheet) and a number of loops interconnecting the sheets (Loris *et al.*, 1998; Vijayan and Chandra, 1999; Barre *et al.*, 2001). The orientation of the anti-parallel strands of the β -pleated sheets changes from one face to the other giving rise to a very strong and rigid

structure (Hamelryck *et al.*, 1998; Van Damme *et al.*, 1998a). The β -sandwich is dome-shaped and the shallow depression at the apex of the dome, made up of four loops, forms the carbohydrate-binding site (Rüdiger, 1997; Hamelryck *et al.*, 1998; Vijayan and Chandra, 1999). Studies by Sharma and Surolia (1997) and by Young and Oomen (1992) demonstrated that variations in the length of a particular loop and/or substitutions of a few amino acids that are involved in sugar-binding activity greatly change the structure of the binding site without affecting the overall three-dimensional structure of the lectin monomer.

Apart from the carbohydrate-binding site, the legume monomer also contains a hydrophobic cavity that consists of mainly conserved hydrophobic residues. Hydrophobic glycosides and hydrophobic derivatives of monosaccharides bind more strongly to this cavity than the analogous non-hydrophobic compounds. This cavity may also be responsible for the non-covalent binding of the plant hormone auxin and adenine (Stroberg *et al.*, 1986; Sharon and Lis, 1990; Van Damme *et al.*, 1998a; Hamelryck *et al.*, 1999).

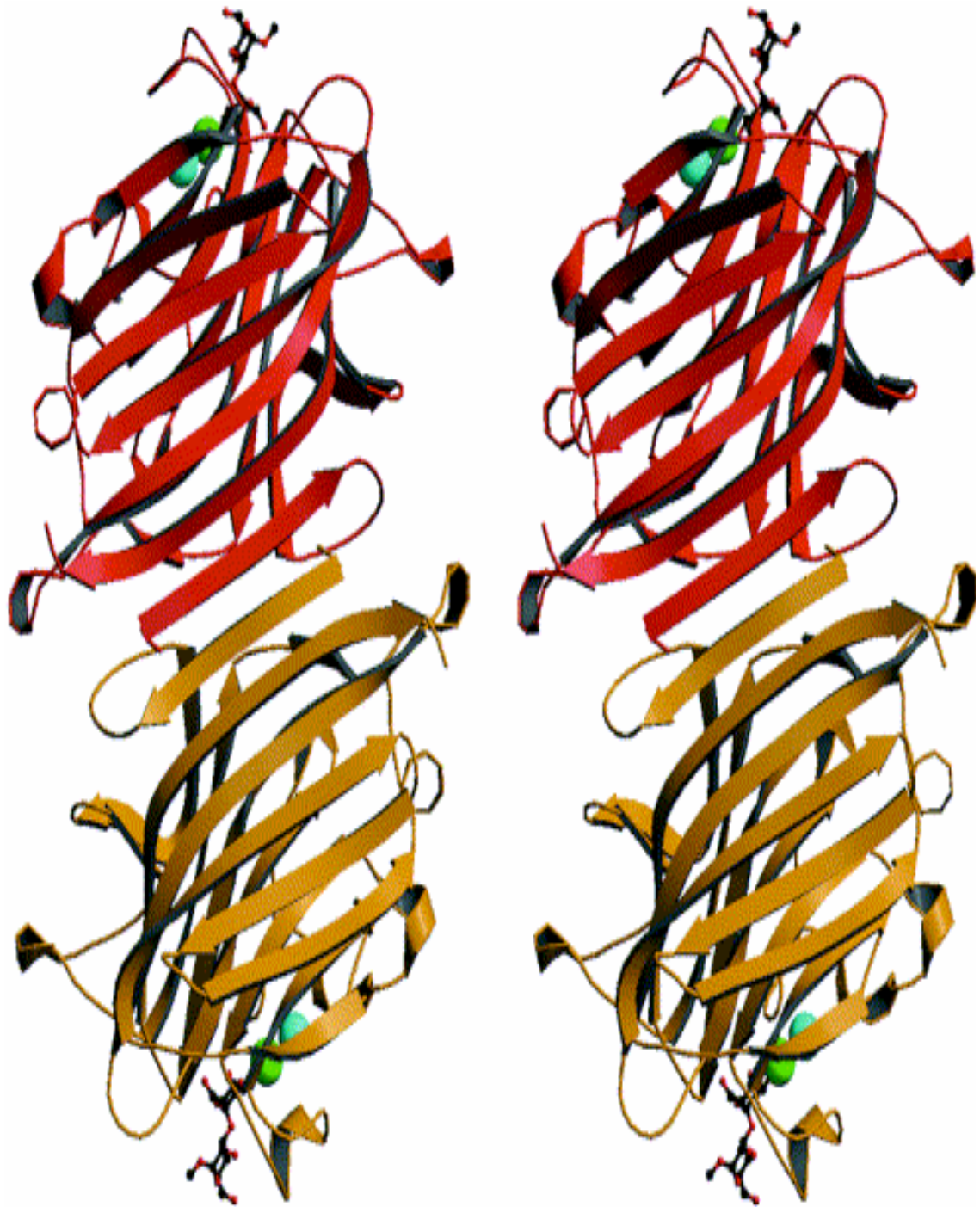


Figure 2.3: Overall structure of the *Pterocarpus angolensis* (mukwa) seed lectin. Schematic representation of the mukwa seed lectin dimer in two orthogonal orientations. One monomer is a darker shade than the other and the spheres represent the calcium and manganese ions. Two bound molecules of Man(1-3)Man are shown in ball and stick representation (taken from Loris *et al.*, 2004).

2.5.3 Lectin-Sugar Interaction

Interaction of lectins with sugar moieties has been studied using a combination of approaches. These approaches include photoaffinity labeling of ligand binding sites and peptide mapping (Díaz-Maurino *et al.*, 1998), computer modeling and ligand docking based on crystallographic studies (Gohier *et al.*, 1996; Neumann *et al.*, 2004), nuclear magnetic resonance (NMR) spectroscopy and molecular modeling (Cheong *et al.*, 1999), site-directed mutagenesis (Sharma *et al.*, 1998) and construction of lectin chimeras (Yamamoto *et al.*, 2000a).

Lectin-sugar interaction occurs in the carbohydrate (sugar)-binding site at the surface of the lectin monomer and is stabilized by hydrogen bonds and hydrophobic interactions (Drickamer, 1997; Calvete *et al.*, 1999). Interactions of the sugar with residues in the combining site are either direct or mediated by water molecules (Weis and Drickamer, 1996). The sugar-binding site is a shallow cleft formed by four distinct regions of polypeptides that form loops connecting the β -strands (Rüdiger, 1997; Sharma and Surolia, 1997; Hamelryck *et al.*, 1998). These loops are designated A, B, C and D (Sharma and Surolia, 1997).

At the centre of the combining site are two highly conserved residues, aspartic acid (Asp) and asparagine (Asn), belonging to loops A and C respectively, which are crucial in sugar binding. These residues, which are positioned by their interactions with the calcium ion, along with the backbone NH of a glycine (Gly) or arginine (Arg) residue (loop B), participate in four key hydrogen bonds with the sugar (Young and Oomen, 1992; Sharma and Surolia, 1997; Rao *et al.*, 1998; Neumann *et al.*, 2004). Thus, for instance, in the mannose/glucose-specific lectins such as concanavalin A and the pea lectin the O ^{δ 1} and O ^{δ 2} of Asp accept hydrogen bonds from C6-OH and C4-OH whereas N ^{δ 2} of Asn and the peptide of Gly or Arg donate to C4-OH and C3-OH respectively. On the other hand, in galactose/N-acetylgalactosamine-specific lectins such as peanut and the soybean agglutinin, O ^{δ 1} and O ^{δ 2} of Asp accepts hydrogen bonds from C4-OH and C3-OH respectively, while N ^{δ 2} of Asn and the

peptide of Gly both donate to C3-OH (Young and Oomen, 1992; Sharma and Surolia, 1997; Hamelryck *et al.*, 1998).

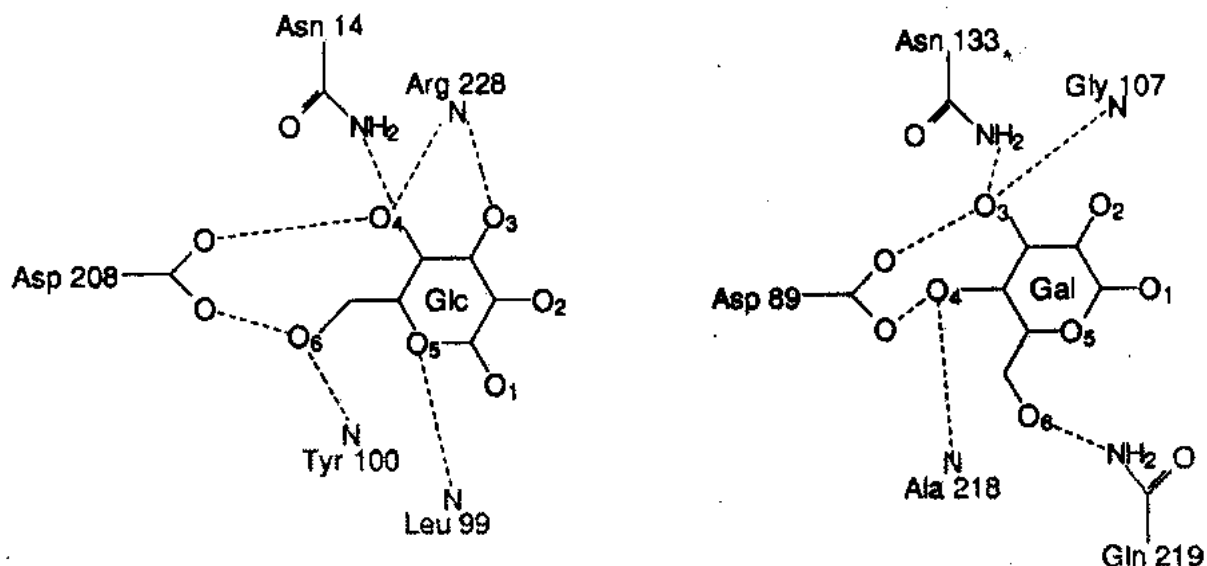


Figure 2.4: Schematic diagram showing monosaccharide recognition by lectins. The left diagram shows the binding of glucose to concanavalin A while the right one shows the binding of galactose to the *Erythrina corallodendron* lectin (taken from Loris *et al.*, 1998).

In addition to the hydrogen bonds, stacking interactions with a hydrophobic residue phenylalanine (Phe), tyrosine (Tyr), tryptophan (Trp) or leucine (Leu) from loop C stabilize the sugar. The basic framework of monosaccharide recognition is completed by interactions with backbone atoms from loop D, which contributes at most three hydrogen bonds (Sharma and Surolia, 1997). The particular residues from loop D, which interact with the sugar, vary from lectin to lectin (Rao *et al.*, 1998).

Different selectivities can be accommodated into this common basic structure because of the shallowness of the sugar-binding site and the limited number of contacts with the sugar (Drickamer, 1997; Sharma and Surolia, 1997).

2.5.4 Sugar-Binding Site

The amino acid residues of the sugar-binding site region vary widely resulting in dissimilar topography of the binding site (Young and Oomen, 1992). However, at the core of the sugar-binding site is a conserved triad, Asp in loop A, Gly or Arg in loop B and Asn in loop C. The site is designed such that a monosaccharide is allowed only in a defined orientation, which depends on the amino acid environment. Firstly an equatorial hydroxyl such as C4-OH in mannose and C3-OH in galactose is directed towards the space in the midst of the triad, that is, O^{δ2} of Asp in loop A, NH of Gly (or Arg) in loop B and N^{δ2} of Asn in loop C. Secondly, an axial hydroxyl or a hydroxymethyl group must be the hydrogen bonding partner for O^{δ1} of Asp in loop A (Sharma and Surolia, 1997).

It is thought that, apart from the primary binding site of the lectin, additional binding in extended or sub-sites occurs. One monosaccharide, usually the terminal one, is bound at the primary binding site of the lectin, with additional monosaccharides along the carbohydrate chain being bound to the secondary sub-sites (Elgavish and Shaanan, 1997; Hamelryck *et al.*, 1998). Interactions made by these sub-sites enhance the affinity for di- and oligosaccharides (Weis and Drickamer, 1996; Loris, 2002). Different lectins with the same monosaccharide specificity may recognize different regions on the surface of oligosaccharides leading to differences in the details of their specificity for oligosaccharides (Iglesias *et al.*, 1982; Loris *et al.*, 2003).

2.5.5 Primary Sugar-Binding Specificity

X-ray crystallography has been employed to assess the structural basis of selective sugar recognition by lectins (Rini, 1995; Weis and Drickamer, 1996; Elgavish and Shaanan, 1997). Selectivity for different sugars results from specific hydrogen bonds to key distinguishing hydroxyl groups. Once an orientation is established, undesired recognition is precluded by steric exclusion of certain groups (Young and Oomen, 1992; Drickamer, 1997).

For instance, the peanut agglutinin binds galactose (Banerjee *et al.*, 1994) but does not bind N-acetyl-galactosamine due to the presence of the glutamic acid residue 129. This residue protrudes into the binding site thereby sterically preventing the accommodation of a bulky substituent at the C-2 position (Sharma *et al.*, 1998).

Generally, monosaccharide components can be classified into two groups based on the recognition pattern of lectins (see Figure 2.4). Group 1 is composed of mannose, glucose and N-acetylglucosamine while group 2 consists of the C-4-epimers galactose and N-acetylgalactosamine. Sugars belonging to group 1 have an equatorial 4-hydroxyl (OH) group whereas group 2 sugars have an axial 4-OH group (Elgavish and Shaanan, 1997; Drickamer, 1997). In their pyranose form, the sugars share the same arrangement of the 3-OH group, the ring oxygen and the carbons at positions 5 and 6. The two groups are distinguished by the orientation of the 4-OH whereas within each group the orientation and substitution at position 2 of the ring varies (Drickamer, 1997).

This primary specificity, however, does not always serve as a clear indication of the overall selectivity, indicating that additional variation in the binding site contributes towards specificity (Shaanan *et al.*, 1991; Rini, 1995; Sharma and Surolia, 1997). For instance the *Griffonia simplicifolia* IV lectin shows no detectable monosaccharide-binding capacity, yet galactose is required for the binding of the Lewis b blood group oligosaccharide (Elgavish and Shaanan, 1997). Multiple binding, mechanisms of additional binding in sub-sites and subunit multivalency possibly also contribute towards the overall selectivity of the bound sugar (Elgavish and Shaanan, 1997; Loris, 2002).

2.5.6 Molecular Basis of Sugar Specificity

The specificity of legume lectins toward different sugars appears to be associated with extensive variations in amino acid residues forming the perimeter of the sugar-binding sites (Shaanan *et al.*, 1991; Young and Oomen, 1992; Sharon and Lis, 1998). The outer hypervariable zone defines the shape of the outer rim of the binding site and may interact with

larger oligosaccharide ligands. The fairly variable zone around the conserved core includes aromatic amino acid residues whose shapes help create monosaccharide specificity (Young and Oomen, 1992). Subtle variations in the shape and size of amino acid residues lining the combining sites possibly could also account for variations in ligand affinity among glucose/mannose-specific lectins (Shaanan *et al.*, 1991). Moreover, the distinct differences in anomeric specificity observed amongst the winged bean lectins I (basic) and II (acidic), the *Erythrina corallodendron* lectin and the peanut agglutinin may be due to the subtle differences in sequence and structure of their D loops (Srinivas *et al.*, 2000).

Comparison of lectin sugar binding sites by Rao and co-workers (1998) showed that the four invariant residues Asp in loop A, Gly in loop B, Asn and an aromatic residue Phe or Tyr, in loop C occupy nearly identical positions. While Asp, Asn and Phe or Tyr constitute the basic binding site itself, the Gly in loop B and one or two amino acid residues in loop D hold the sugar tightly on either side of the site by means of hydrogen bonds (Rao *et al.*, 1998). Given that the Gly in loop B is invariant, variations in conformation and sequence of loop D therefore possibly control sugar specificity in these lectins (Rao *et al.*, 1998).

Sharma and Surolia (1997), however, suggested that the size of the sugar-binding loops is essential in determining sugar specificity (see Figure 2.5). The highly variable loop D, which has long been recognized as an important discriminatory factor of the particular sugar bound by a given legume lectin (Elgavish and Shaanan, 1997; Sharma and Surolia, 1997; Hamelryck *et al.*, 1998; Loris *et al.*, 2003), shows gross difference in size between mannose/glucose-specific and galactose/N-acetylgalactosamine-specific lectins (Sharma and Surolia, 1997). While there is significant sequence variation of loop D in all the mannose/glucose-specific legume lectins analyzed by Sharma and Surolia (1997), the loop is identical in size (Thomas and Surolia, 2000). The galactose/N-acetylgalactosamine-specific lectins characteristically have longer D loops, whilst amongst known chitobiose or N-acetylglucosamine-specific lectins, the size of loop D is conserved in spite of the extensive sequence differences. The size of the backbone of loop D has therefore been implicated as the

primary determinant of sugar specificity. Sequence analyses, however, show that sequence differences and variation in the size of loop C, which further influences the size and openness of the binding site, also contribute to determining sugar specificities (Sharma and Surolia, 1997; Loris, 2002; Loris *et al.*, 2004).

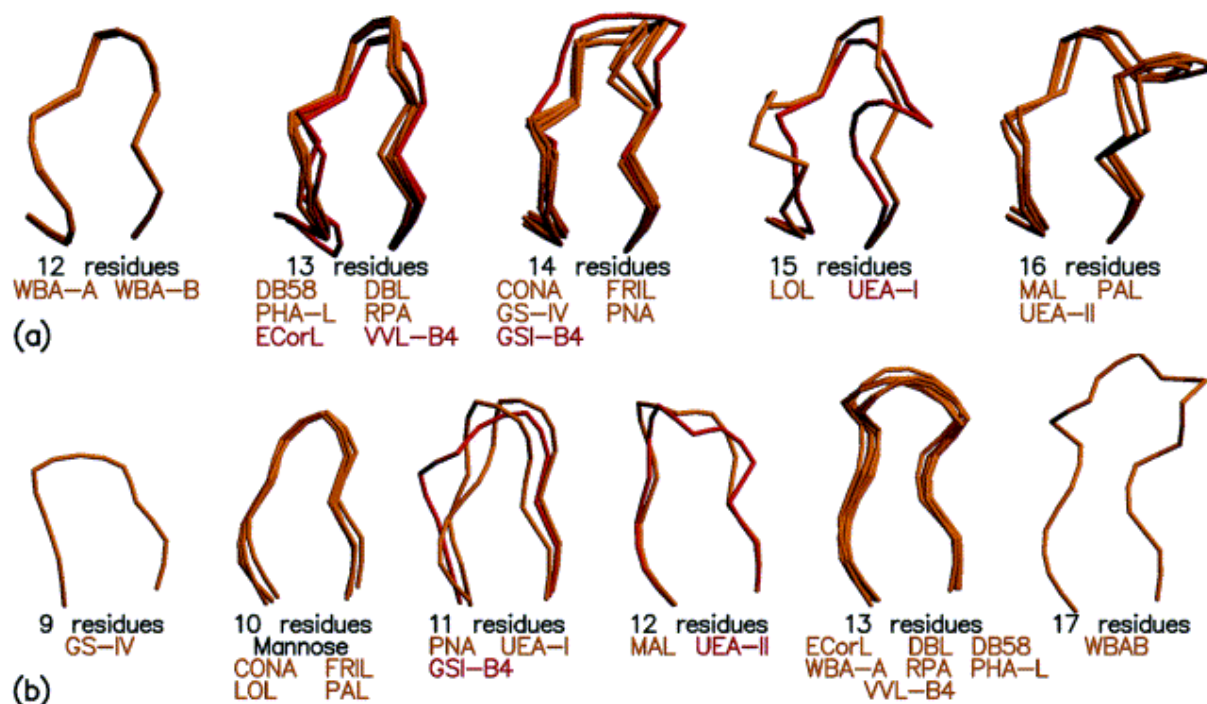


Figure 2.5: Loop conformations in the sugar-binding sites of legume lectins. (a) conformations observed for loop C and (b) conformations observed for loop D suggesting the use of the canonical loop conformations to modulate sugar specificity (taken from Loris *et al.*, 2004). (see Appendix B for abbreviations).

In principle, using site directed mutagenesis, it should be possible to engineer lectins with desired specificities by changing the size of the variable loops (Sharma and Surolia, 1997; Loris, 2002). However, in all such cases it would be imperative to retain the conserved residues in loops A (Asp), B (Gly), and C (Asn and Phe or Tyr) (Van Eijsden *et al.*, 1992; Adar and Sharon, 1996; Nishiguchi *et al.*, 1997). Loop D, which constitutes the more variable edge of the binding pocket, would be targeted for changing the primary specificity by varying both its size and sequence (Sharma and Surolia, 1997; Rao *et al.*, 1998).

2.6 *Pterocarpus* Genus

The genus *Pterocarpus* belongs to the family Leguminosae, subfamily Papilionoideae, tribe Dalbergieae. The term *Pterocarpus*, derived from Greek, means ‘winged fruit’ and describes the unusual seed pods which consist of a central hardened seed case surrounded by a broad membranous wing (Rojo, 1977; Palgrave, 1981). About 30 species belonging to this genus have been described and these are distributed throughout the tropics. Most of the species originated and are distributed in Africa and Asia with some such as *P. indicus* being found widespread in South and Central America (Echemendia-Blanco, 2001).

All *Pterocarpus* species are very attractive trees with heights of up to 33 metres. Species can be distinguished from each other largely by the size of their fruit although in some cases, flowers and rarely leaves are used (Rojo, 1977).

2.7 *Pterocarpus angolensis*

Pterocarpus angolensis (*P. angolensis*), known as the ‘mukwa’ tree in Zimbabwe and as ‘kiaat’ in South Africa, has brown, spiky seed pods which remain on the tree long after the leaves have fallen off. The specific name, ‘*angolensis*’, means ‘of Angola’, and the brilliant red, sticky substance that the bark exudes when injured, gives it the common name ‘bloodwood’. The tree is also referred to as ‘wild teak’ as a result of the resemblance of its wood to the unrelated teak, *Tectona grandis*, from tropical East Asia (Palgrave, 1981; Funston, 1993).

2.7.1 Distribution

The mukwa tree is indigenous to East Africa, extending from Tanzania to the northern provinces of South Africa. Across the continent, it is found in Angola through Zambia, the greater part of Zimbabwe and in Mozambique (Schwartz *et al.*, 2002). Populations of this tree are denser on the Kalahari sands as the tree has a preference for well-drained, deep sandy soil

(Palgrave, 1981; Van Wyk, 1993). These wild populations are diminishing, yet the tree is not common in cultivation due to difficulty with seed germination (Shackleton, 2002).

2.7.2 Description

Pterocarpus angolensis (mukwa) (Figure 2.6), is a deciduous, wide-spreading tree with a slightly flattened crown. The tree reaches about 15 metres in height and has a rough, dark brown bark. The leaves have 5 to 9 pairs of obovate leaflets that are shiny, intensive green in colour and characteristically hang downwards. Sweetly scented orange-yellow flowers emerge in spring immediately prior to, or with, the new leaves. The circular fruit turns from green to light brown when mature. The central part of the fruit, which normally contains only one kidney-shaped seed, is densely covered with harsh bristles. The seed has a thin, hard red-brown coat. The wood is two-coloured, the softer yellowish sapwood and the hard reddish-brown heartwood (Palgrave, 1981; Van Wyk and Van Wyk, 1997).

2.7.3 Uses

Pterocarpus angolensis is widely exploited for its timber, which is durable, medium hard and very easy to work. The wood is utilized for furniture, veneer, curios, boat-construction and as general purpose timber (Van Wyk, 1993; Schwartz *et al.*, 2002).

Different parts of the tree are used to treat numerous diseases throughout Africa. The root is believed to cure malaria, gonorrhoea and black water fever. The bark is used to treat headaches, stomach-aches, mouth sores and rashes. The sap is used for eye treatment against cataracts as well as treatment of urinary schistosomiasis, ringworms and ulcers (Palgrave, 1981; Schwartz *et al.*, 2002). Traditionally, the red sap is used as a dye and in some areas it is dried, pounded and mixed with animal fat to make cosmetics (Palgrave, 1981).

Being a legume, the mukwa tree is nitrogen-fixing and is used for soil conservation. Leaves and shoots are used as fodder and the scented flowers make the tree a good bee plant (Palgrave, 1981).

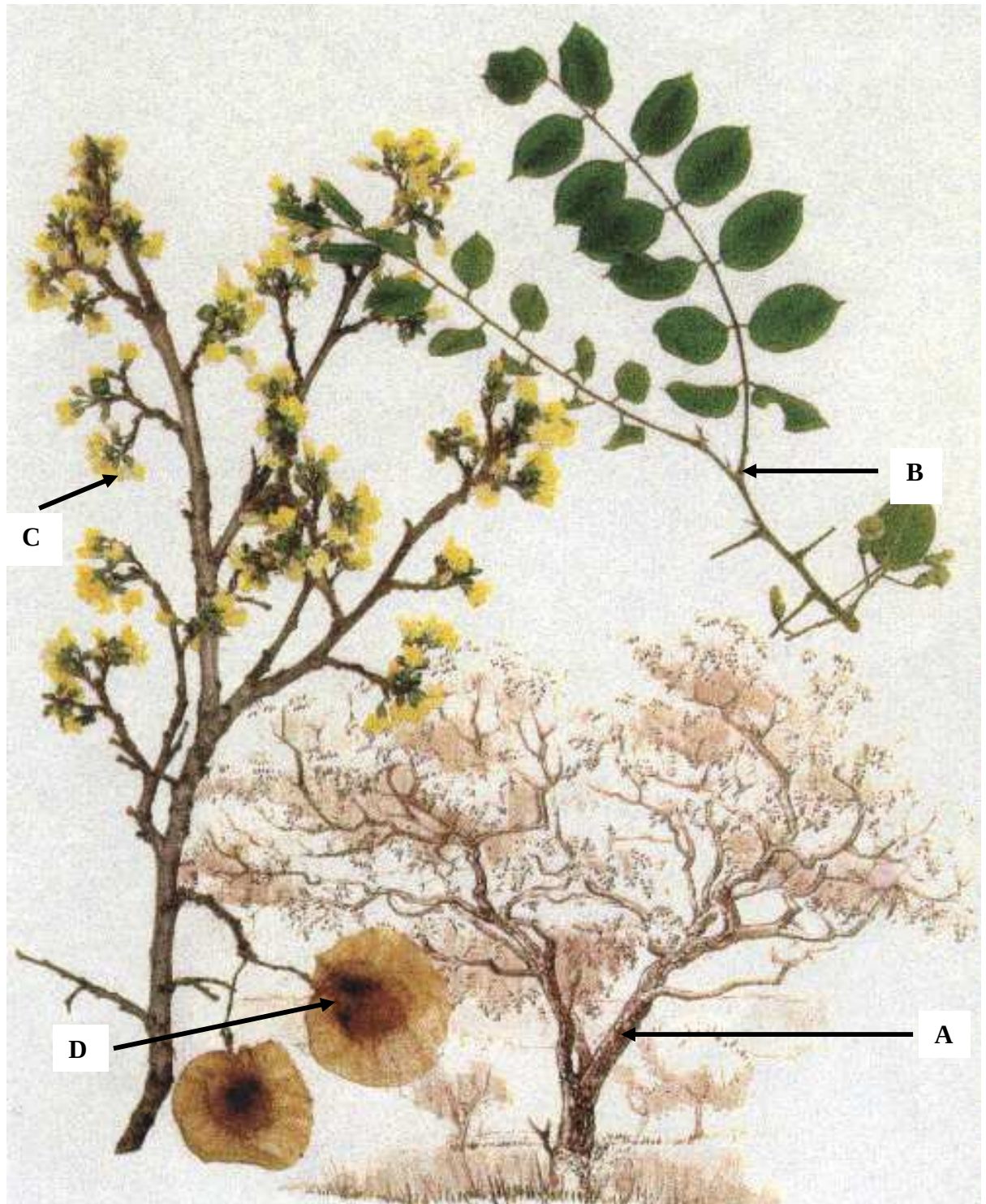


Figure 2.6: Picture of *Pterocarpus angolensis* (mukwa tree). A shows the tree, B the leaves, C the flowers and D the fruit (adapted from Palgrave, 1981).

2.7.4 Lectin Activity

A study of haemagglutinins of trees and shrubs of Zimbabwe showed the presence of lectin activity in the seeds of the mukwa tree. This lectin haemagglutinated only protease-treated human cord erythrocytes and not adult erythrocytes, hence it was concluded that the lectin was specific for i-antigens (Moore, 1979). Kaul and co-workers (1991) showed that this seed lectin binds glycoproteins such as ovalbumin, horseradish peroxidase and porcine mucin. Lectin activity has since been detected in other tissues such as the roots, bark, stem as well as the leaves of this tree (I. Ncube, personal communication).

The lectin from mukwa seeds has been purified and extensively characterized (Maramba, 1998; Loris *et al.*, 2003). This seed lectin is a homodimer made up of 28 kDa subunits that are not held together by covalent linkage. The protein has virtually no α -helix structure and is mainly folded into β -pleated sheet segments. The metal cations Mn^{2+} and Ca^{2+} are more tightly bound to this lectin than generally is the case for other legume lectins and the lectin is stable over a wide range of pH values, 3 to 10, (S. Beeckmans, unpublished).

The mukwa seed lectin, which agglutinates rabbit erythrocytes, has a classical mannose-specificity loop yet its metal-binding loop (loop C) resembles that of *Ulex europaeus* and *Maackia amurensis*, lectins of unrelated specificity. Affinity for $Man(\alpha 1-2)Man$ and $Man(\alpha 1-3)[Man(\alpha 1-6)]Man$ by the mukwa lectin is very high although its fine specificity differs from that of other known mannose/glucose-specific lectins (Loris *et al.*, 2004).

3. MATERIALS AND METHODS

3.1 MATERIALS

A pUC18 vector carrying the cDNA encoding the wild type mukwa seed lectin, clone Muk151QII28, was provided by Professor H. De Greve, Institute of Molecular Biology and Biotechnology/ VUB (Brussels, Belgium). The cloned cDNA consists of part of the signal peptide, the mature lectin, the C-terminal peptide and the 3' untranslated region.

*Ex Taq*TM DNA polymerase was obtained from Takara Biomedicals (Japan).

Oligonucleotide primers, used for PCR, sequencing and for *in vitro* mutagenesis, were designed by Professor H. De Greve, Institute of Molecular Biology/ VUB (Brussels, Belgium) and ordered from Invitrogen Life Technologies (USA).

The QIAprep Spin Miniprep kit for plasmid DNA isolation and the PCR Purification kit for purification of PCR fragments, were obtained from Qiagen (Germany).

The Thermo Sequenase Radiolabelled Terminator Cycle Sequencing kit was obtained from United States Biochemical Corporation (USA). X-ray film and developer were from Fuji (Japan) and AGFA-GEVAERT (Germany) respectively. The PC/GENE and Genetics Computer Group (GCG) sequence analysis software packages were used for nucleotide and amino acid sequence analysis.

The Unique Site Elimination Mutagenesis kit was obtained from Amersham Biosciences (UK).

Rabbit antibodies directed against the mukwa seed lectin were initially provided by Professor S. Beeckmans, Institute of Molecular Biology and Biotechnology/ VUB (Brussels, Belgium).

Rabbit blood was from the Veterinary and Agrochemical Research Centre, Brussels, Belgium or from the Animal House, University of Zimbabwe, and human blood from the Brussels University Hospital Blood Transfusion Centre, Belgium.

Reagents for preparation of bacterial growth media were obtained from Difco Laboratories (USA).

Klenow fragment, T4 DNA polymerase, and T4 DNA ligase were obtained from MBI Fermentas (Germany).

Restriction enzymes were obtained from MBI Fermentas (Germany); Promega (USA); New England Biolabs (USA); Amersham Biosciences (UK) and Boehringer Mannheim (Germany). These were used according to the manufacturer's recommendations.

Extravidin phosphatase, protein A-biotin, anti-rabbit IgG phosphatase and protein molecular mass standards, phosphorylase B (92 kDa), catalase (60 kDa), fumerase (50 kDa), aldolase (40 kDa), carbonic anhydrase (30 kDa), trypsin inhibitor (20 kDa) and lysozyme (14.3 kDa), were obtained from Sigma Chemical Company (UK).

The bicinchoninic acid (BCA) kit was obtained from Pierce Chemical Company (USA).

Porcine and bovine plasma powders were obtained from Vepro Products (Belgium).

Lambda DNA used as molecular mass marker was obtained from MBI Fermentas (Germany).

The antibiotics carbenicillin and chloramphenicol were from DUCHEFA (The Netherlands).

All the other chemicals not available locally were obtained from Sigma Chemical Company (UK) or Merck (Germany).

For experiments involving DNA, prior to use, Eppendorf tubes, pipette tips and all other plasticware were autoclaved at 15 lb/in², 121°C for 20 min and sterile water was used for making all solutions.

3.1.1 Plasmids and Bacterial Strains

Plasmids pUC18, BL21-CodonPlus-RIL and pBADMycHisA were obtained from MBI Fermentas (Germany), Stratagene (USA) and Invitrogen Life Technologies (USA) respectively.

The pUC18 plasmid (Yanisch-Perron *et al.*, 1985), contains the β -lactamase gene which confers resistance to ampicillin or carbenicillin and the origin of replication of pBR322. The plasmid also contains the α -peptide of the β -galactosidase (lac Z) gene, and a multiple cloning site of one of the M13 sequencing vectors. It differs with pUC19 only in the orientation of the multiple cloning site (see Appendix E). In pUC18 the *Hind* III recognition site is closest to the M13/pUC Forward Sequencing Primer binding site (Messing, 1979).

Plasmid BL21-CodonPlus-RIL contains extra copies of the genes that encode tRNAs for the arginine codons AGA and AGG, the isoleucine codon AUA and the leucine codon CUA, which are rarely used in *Escherichia coli* (*E. coli*). Transformation of *E. coli* with this plasmid therefore enhances the expression of these codons (Stratagene, 2000).

The expression vector pBADMycHisA, 4 096 bp in size, has the *araBAD* promoter (P_{BAD}), which provides tight, dose-dependent regulation of heterologous gene expression (Guzman *et al.*, 1995). This vector has an optimized ribosome binding site; an *rrnB* transcription termination region; an *araC* gene, which encodes the regulatory protein for tight regulation of the P_{BAD} ; and an initiator ATG, which provides a translational initiation site for the protein to be expressed. It also contains the β -lactamase gene, a pUC-derived origin, a multiple cloning site and has a C-terminal *myc* epitope followed by a polyhistidine tag (Invitrogen Life Technologies, 2000) (see Appendices F and G).

The *E. coli* strain K514 was used as the host to propagate the plasmid pUC18. The genotype of this strain is *sup E44 thr-1 leuB6 thi-1 lacY1 ton A21 hsdR* (Colson *et al.*, 1965).

The *E. coli* TOP10 strain (Invitrogen Life Technologies, 2000) was used as host for the expression of the mukwa seed lectin constructs. It has the genotype *F⁺mcrA Δ (mrr-hsdRMS-*

mcrBC) $\Phi 80$ *lacZ* Δ M15 Δ *lacX*74 *deoR* *recA*1 *araD*139 Δ (*araA-leu*)7697 *galU* *galK* *rpsL* *endA*1 *nupG*. This strain is able to transport L-arabinose but cannot metabolize it.

The *E. coli* strain WK6*mutS*, deficient in DNA repair mechanism, was used as the host for transformation with plasmid after *in vitro* mutagenesis. This strain has the genotype Δ (*lac-proAB*) *galE* *strA* *mutS*215::Tn10F'(*lacI*^q *lacZ* Δ M15 *proA*⁺*B*⁺) (H. De Greve, personal communication).

3.2 METHODS

Standard molecular cloning laboratory techniques (Sambrook and Russell, 2001) were followed with modifications in some cases.

3.2.1 Amplification of the cDNA Encoding the Mature Wild Type Mukwa Seed Lectin By PCR

The polymerase chain reaction (PCR) technique was used to amplify the cDNA encoding the mature wild type mukwa seed lectin. Repetitive cycles of DNA denaturation, primer annealing and extension enable one to adequately amplify the appropriate target DNA segment to levels that are detectable by ethidium bromide staining of agarose gels.

The reaction requires the DNA template containing the target sequence to be amplified, two oligonucleotide primers, four deoxyribonucleoside triphosphate substrates (dATP, dCTP, dGTP and dTTP) and a thermostable DNA polymerase such as *Thermus aquaticus* (*Taq*) DNA polymerase in a buffered solution.

3.2.1.1 Primer Design

Three specific oligonucleotide primers were designed based on the already sequenced clone Muk151QII28, pUC18 vector carrying the cDNA encoding the wild type mukwa seed lectin (GenBank Accession Number AJ426056). The cDNA, in clone Muk151QII28, encodes

part of the signal peptide, the lectin till residue A²⁴¹ (which will be called hereafter ‘the mature mukwa seed lectin’), the amino acid residue Q²⁴², the C-terminal peptide and the 3’ untranslated region. The specific primers, designated Muk-23, Muk-24 and Muk-25 (see Figure 3.1), were ordered from Invitrogen Life Technologies (USA).

X-ray crystallography had shown that the first amino acid of the mature mukwa seed lectin, glutamine, is in a pocket in the interior of the protein. It was assumed that addition of the initiator methionine from the expression vector probably would interfere with proper protein folding. Leaving out the first amino acid of the mature lectin could conceivably accommodate the methionine and hopefully still have the protein in the correct conformation. For this reason, two N-terminal primers, Muk-23 and Muk-24, were designed right at the beginning. Primers Muk-23 and Muk-24 have sequences

5’-CAAGATTCCCTTTCCTTCGGCTTC-3’ and

5’-GATTCCCTTTCCTTCGGCTTCCTA-3’ respectively. Muk-23 allows for amplification of the mature mukwa seed lectin-coding region from the first amino acid residue whereas Muk-24 amplifies from the second amino acid residue, aspartic acid.

The Muk-25 primer, which has the sequence

5’-GGGAATT**C**TAAGCAGTGTAACAAGGTTGAGGT-3’, includes an *EcoRI* restriction endonuclease recognition site, shown in bold, and a stop codon, underlined. Amplification with Muk-25 introduced a stop codon just after alanine, the last amino acid residue of the mature mukwa seed lectin that could be identified from the crystal structure. Incorporation of the stop codon meant that translation of the lectin would be terminated without expressing the amino acid residue Q²⁴² and the C-terminal peptide. The amino acid residue Q²⁴² and the C-terminal peptide, residues K²⁴³ to M²⁵² could not be seen in the crystal structure. The *EcoRI* recognition site was incorporated to allow for directional cloning of the PCR-generated fragments.

Muk151QII28	-----M L L N K A Y S -----ATGCTACTGAACAAAGCATACTC	23
	a b ↓ ↓ Q D S L S F G F P T F P S D Q K	
1 Muk151QII28	CCAAGATTCCCTTTTCCTTCGGCTTCCCTACGTTTCCTTCAGACCAGAAAA → Muk-23 and Muk-24	73
17 Muk151QII28	N L I F Q G D A Q I K N N A V Q L ACCTAATCTTCCAAGGTGATGCACAGATTAAAAACAACGCAGTCCAGCTT	123
34 Muk151QII28	T K T D S N G N P V A S T V G R I ACCAAGACAGACAGTAATGGTAACCCAGTGGCAAGCACTGTTGGGCGAAT	173
51 Muk151QII28	L F S A Q V H L W E K S S S R V CTTATTCTCGGCGCAAGTGCACCTTTGGGAAAAAAGCTCAAGCAGAGTGG	223
67 Muk151QII28	A N F Q S Q F S F S L K S P L S N CAAACCTTTCAATCCCAGTTCAGCTTTTCCCTCAAATCACCGCTTTCCAAT	273
84 Muk151QII28	G A D G I A F F I A P P D T T I P GGAGCCGACGGCATTGCCTTCTTCATCGCACCGCCGGATACTACCATACC	323
101 Muk151QII28	S G S G G G L L G L F A P G T A GAGCGGTTTCGGGCGGAGGTCTCCTAGGGCTCTTTGCACCTGGAAGTCTC	373
117 Muk151QII28	Q N T S A N Q V I A V E F D T F Y AAAACACGTCCGCGAACCAAGTGATTGCGGTTGAGTTTGACACCTTCTAT	423
134 Muk151QII28	A Q D S N T W D P N Y P H I G I D GCTCAAGACTCTAACACTTGGGATCCGAATTACCCGCACATTGGAATCGA	473
151 Muk151QII28	V N S I R S V K T V K W D R R D TGTCAACTCCATTAGGTCTGTGAAGACTGTGAAATGGGATAGGAGAGATG	523
167 Muk151QII28	G Q S L N V L V T F N P S T R N L GTCAATCCCTTAATGTCTCGTAACATTTAACCCTAGCACTCGAAACCTG	573
184 Muk151QII28	D V V A T Y S D G T R Y E V S Y E GATGTCGTTGCTACTTACTCTGATGGTACAAGATACGAGGTGTCTTATGA	623
201 Muk151QII28	V D V R S V L P E W V R V G F S GGTTGACGTGAGGTCCGTGCTTCCAGAATGGGTTAGAGTTGGATTCTCAG	673
217 Muk151QII28	A A S G E Q Y Q T H T L E S W S F CTGCCTCAGGAGAGCAATATCAAACGCATACTCTTGAATCATGGTCTTTC	723
234 Muk151QII28	T S T L L Y T A Q K K G E N L A L ACCTCAACCTTGTTGTACACTGCTCAGAAGAAGGGCGAGAATCTTGCACT	773
	← TAGAATTCC Muk-25	
251 Muk151QII28	E M TGAGATGTGATCTGGAGTGATTTCTATGTCATATCCATCATGTATGATAT	823

Figure 3.1: Primer design based on clone Muk151QII28, a pUC18 vector carrying the cDNA encoding the wild type mukwa seed lectin. The black arrows, a and b, pointing downwards indicate the 5' ends of primers Muk-23 and Muk-24 respectively while the horizontal arrow shows where Muk-25 binds (adapted from GenBank Accession Number AJ426056).

3.2.1.2 Polymerase Chain Reactions (PCR)

The polymerase chain reactions were carried out in a total volume of 50 μ l. Clone Muk151QII28, used as template DNA, was diluted serially from 1: 10 to 1: 10 000. One microlitre of each of the 1: 100, 1: 1 000 and 1: 10 000 template DNA dilutions was diluted further to 10 μ l. Five microlitres of 10 x *Ex Taq*TM polymerase reaction buffer, 4 μ l of 2.5 mM dNTPs, 1 μ l (of a 20 μ M solution) of each of the two primers, 28.75 μ l of water, 0.25 μ l (1.25 units) of *Ex Taq*TM DNA polymerase and 10 μ l of the template DNA were mixed together for each reaction. The full composition of the 10 x *Ex Taq*TM polymerase reaction buffer is proprietary but the magnesium concentration is 20 mM (Takara Biomedicals, 2000). Reactions were made up as master mixes to minimize pipetting errors. Negative control reactions without template DNA were always included.

An automated PCR protocol was used routinely in a programmable thermocycler (Perkin Elmer GeneAmp PCR System 9600, USA). The template DNA was initially denatured at 94°C for 50 sec. The second step of the program consisted of 10 cycles of template denaturation at 94°C for 10 sec, primer annealing at 55°C for 30 sec and primer extension at 68°C for 60 sec. These 10 cycles were followed by 20 cycles of template DNA denaturation at 94°C for 10 sec, primer annealing at 55°C for 30 sec and primer extension at 68°C for 60 sec with 15 sec increments with each cycle. Finally, at the end of all cycles, maintaining the samples at 72°C for 10 min ensured that all the PCR fragments were fully extended.

A 5 μ l aliquot of each PCR product was analyzed by resolution on a 0.8% (w/v) analytical agarose gel in TBE buffer (89 mM Tris base, 89 mM boric acid and 2 mM Na₂EDTA, pH 8.0). The gel was run at a constant voltage (80 V) and the bands visualized by staining in 0.5 μ g/ml ethidium bromide and viewing by ultra-violet (302 nm) illumination on a transilluminator.

3.2.1.3 Purification of PCR Fragments

Upon completion of PCR, excess primers, nucleotides, *Taq* DNA polymerase and salts must be removed from the amplification products before the latter can be cloned. Removal of the contaminants was done using the PCR purification kit (Qiagen, 2000b).

The PCR products generated using the same primer pair were pooled together in an Eppendorf tube and mixed with 5 volumes of the Binding buffer PB. The mixture was applied onto a QIAquick spin column and centrifuged for 1 min at 13 000 rpm in a microfuge to allow the DNA to bind to the column. The flow-through was discarded and the column washed with 750 µl of Wash buffer PE by centrifuging at 13 000 rpm for 1 min. The flow-through was discarded and residual Wash buffer PE removed by centrifuging again at 13 000 rpm for 2 min. The QIAquick column was placed on a new Eppendorf tube and 50 µl of Elution buffer EB, pre-warmed to 65°C, applied onto the column. After incubating at room temperature for 10 min, the DNA was eluted by centrifuging at 13 000 for 1 min.

Five microlitres of each sample were analyzed by resolution on a 0.8% (w/v) agarose gel to check for DNA recovery, quality and to estimate quantity.

3.2.1 Cloning of the Amplified cDNA Encoding the Mature Wild Type Mukwa Seed Lectin into the Expression Vector pBADMycHisA

Cloning of PCR-amplified products directly by blunt-end ligation is usually unsuccessful. The failure is most probably due to alterations of the DNA termini, which renders them incompatible for ligation into vectors. For instance, *Taq* DNA polymerase adds 3'-protruding nucleotides, a single A in 75% of the cases, in a template-independent manner (Hitti and Bertino, 1994).

Incorporation of an *EcoRI* restriction endonuclease recognition site into the PCR primer Muk-25 meant that the 3' ragged ends of the PCR products could be eliminated by digestion with *EcoRI*. Removal of the ragged ends would then facilitate directional cloning into the compatible *EcoRI* site in the polylinker of the pBADMycHisA vector.

3.2.2.1 Digestion of pBADMycHisA with the Restriction Endonuclease *NcoI*

The expression vector pBADMycHisA has an *NcoI* restriction endonuclease recognition site at the 5' end of its multiple cloning site. Digesting this vector DNA with *NcoI* therefore linearized it. The single digest was carried out in a total volume of 20 µl. Fifteen microlitres (approximately 750 ng) of vector DNA, 2 µl of 10 x NEB buffer 4, 2 µl of water and 1 µl (10 units) of *NcoI* were mixed together and incubated at 37°C for 2 hr. The 10 x NEB buffer 4 was composed of 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate and 1 mM DTT, (pH 7.9 at 25°C).

After digestion, a 3 µl aliquot of the *NcoI*-digested pBADMycHisA was run on a 0.8% (w/v) agarose gel to check for complete digestion.

3.2.2.2 Filling-In of the *NcoI*-digested pBADMycHisA

Digestion of pBADMycHisA with *NcoI* generated a cohesive end that was then filled in with T4 DNA polymerase. This filling-in reaction then provided an initiator ATG in the expression vector.

The filling-in reaction was carried out in a total volume of 20 µl. To 17 µl of the *NcoI*-digested pBADMycHisA, 2 µl of 2.5 mM dNTPs and 1 µl (5 units) of T4 DNA polymerase were added and mixed well. No additional buffer was included as the buffer from the *NcoI* digestion was sufficient for this reaction as well. The reaction was allowed to proceed for 15 min at 12°C after which the T4 DNA polymerase was heat-inactivated by incubating at 65°C for 10 min.

3.2.2.3 Digestion of the *NcoI*-filled-in pBADMycHisA with *EcoRI*

Digestion of the expression vector with a second restriction endonuclease, *EcoRI*, would generate a cohesive end that would be compatible with the PCR-amplified fragments.

The *EcoRI* digestion was carried out in a total volume of 25 μ l. The 20 μ l of *NcoI*-filled-in vector DNA were mixed with 2.5 μ l of 10 x high salt buffer (100 mM Tris-HCl, 50 mM NaCl and 10 mM $MgCl_2$), 1.5 μ l of water and 1 μ l (10 units) of *EcoRI*. The reaction mixture was incubated at 37°C for 2 hr, after which a 3 μ l aliquot was run on a 0.8% (w/v) agarose gel to check for quality and to estimate the quantity of DNA.

3.2.2.4 Polishing of the PCR Fragments

Ex TaqTM DNA polymerase usually adds an A at the 3'-end of the PCR-generated fragment (Takara Biomedicals, 2000). This was removed using the 3' to 5' exonuclease activity of Klenow fragment.

The polishing reaction was carried out in a total volume of 20 μ l. Ten microlitres (approximately 400 ng) of PCR DNA were mixed with 2 μ l of 10 x Klenow fragment buffer, 2 μ l of 0.5 mM dNTP mix, 1 μ l (10 units) of Klenow fragment, 1 μ l of 100 mM ATP, and 4 μ l of water and incubated at 37°C for 30 min. The 10 x Klenow fragment reaction buffer consisted of 500 mM Tris-HCl (pH 8.0 at 25°C), 50 mM $MgCl_2$ and 10 mM DTT.

At the end of the incubation period, the PCR DNA was purified using the PCR purification kit (Qiagen, 2000b) and eluted in 30 μ l Elution buffer EB. A 3 μ l aliquot was run on a 0.8% (w/v) agarose gel to check for DNA recovery and quality.

3.2.2.5 Digestion of the PCR-Generated Fragments with *EcoRI*

The PCR primer Muk-25 had an *EcoRI* recognition site incorporated into it and therefore cohesive 3' ends could be generated in all PCR-generated fragments by digesting them with *EcoRI*.

Digestion was carried out in a total volume of 20 μ l. To 17 μ l of the polished PCR DNA, 2 μ l of high salt buffer and 1 μ l (10 units) of *EcoRI* were added and the mixture

incubated at 37°C for 2 hr. A 3 µl aliquot was run on 0.8% (w/v) agarose gel to check for DNA quality and to estimate quantity.

3.2.2.6 Ligation of pBADMycHisA Vector to PCR-Generated Fragments

Joining of the vector and the insert (PCR-generated fragments) was catalyzed by T4 DNA ligase. This enzyme catalyzes the formation of phosphodiester bonds, in the presence of ATP, between double-stranded DNAs with 3' hydroxyl and 5' phosphate termini.

The ligation reaction was carried out in a total volume of 20 µl. Nine microlitres of the *EcoRI-NcoI*-filled-in-pBADMycHisA vector DNA, 9 µl of *EcoRI*-PCR insert DNA, 2 µl of 10 x T4 DNA ligase buffer and 1 µl (5 units) of T4 DNA ligase were mixed together and incubated at 16°C overnight. The 10 x T4 DNA ligase buffer contained 200 mM Tris-HCl, pH 7.6, 100 mM MgCl₂, 5 mM ATP, pH 7.4 and 100 mM DTT.

3.2.2.7 Preparation of Competent Cells

Bacteria in the mid-log phase of growth are capable of taking up exogenous DNA. It has been shown that if the host bacteria are first treated with CaCl₂ solution, their uptake of foreign DNA is enhanced.

A 15% glycerol stock of *E. coli* K514 cells (stored at -80°C) was thawed on ice and a loopful of the cells aseptically streaked onto Luria-Bertani (LB) agar (10 g Bacto-tryptone, 5 g Bacto-yeast extract, 10 g NaCl and 15 g Bacto-agar per litre, pH 7.5). The plate was incubated at 37°C overnight.

Single colonies were aseptically picked, inoculated into 5 ml aliquots of LB broth (without the agar) and grown overnight with shaking at 250 rpm (New Brunswick Innova Shaker, USA). One millilitre of the overnight culture was inoculated into 100 ml of LB liquid broth in a 500 ml flask and grown at 37°C with shaking at 250 rpm to an absorbance of 0.40-0.45 at 660 nm. The cell suspension was then chilled on ice for 30 min, centrifuged at 4 000 x

g for 15 min at 4°C (Sorvall RC-5B Refrigerated Superspeed Centrifuge, Du Pont Instruments, USA) and the supernatant discarded.

The cells were re-suspended in half the original culture volume (50 ml) of an ice-cold sterile solution of 50 mM CaCl₂ and 10 mM Tris-HCl, pH 8.0. The cell suspension was incubated on ice for 15 min, centrifuged at 4 000 x g for 15 min at 4°C and the supernatant discarded. The cells were re-suspended in one-tenth the original volume (10 ml) of the ice-cold calcium chloride solution. Sterile glycerol was added to a final concentration of 15%. Two hundred microlitre aliquots were dispensed into pre-chilled Eppendorf tubes and stored at -80°C. Competence was assessed by transforming these cells with pUC18 and plating on LB agar supplemented with 100 µg/ml carbenicillin. Untransformed cells were also plated on LB agar plates containing 100 µg/ml carbenicillin as a negative control.

3.2.2.8 Transformation of Competent Cells by the CaCl₂ Method

The protocol given by Hanahan and co-workers (1991) was used to transform competent *E. coli* K514 cells.

Ten microlitres of the ligated plasmid vector were added to 200 µl competent cells, the Eppendorf tube tapped and incubated on ice for 10 min followed by incubation at 37°C for 5 min. The cell suspension was then added to 2 ml of LB broth in a 17 x 100 mm polypropylene tube and incubated at 37°C for 2 hr with shaking at 210 rpm. One hundred microlitres of the cell suspension were then plated out on LB agar plates containing 100 µg/ml carbenicillin and incubated at 37°C overnight.

3.2.2.9 Colony Screening PCR for Possible Recombinants

Only those bacteria carrying plasmid can grow on agar plates supplemented with carbenicillin. These bacteria are resistant to carbenicillin because the plasmid carries the *bla* gene that codes for β-lactamase, a periplasmic enzyme, which cleaves the β-lactam ring of the

antibiotic. Once the ring has been cleaved, the antibiotic no longer interferes with the terminal reaction in the bacterial wall synthesis.

Thirty colonies from each transformation reaction were screened for possible recombinants. A single colony of cells was picked using the tip of a sterile toothpick and the tip used to inoculate a gridded LB agar plate containing 100 µg/ml carbenicillin. The same tip was then dipped into 50 µl of sterile water, into which five colonies were pooled together. One microlitre of each pool was made up to 10 µl with water and used as template DNA for the PCR reaction. The PCR reactions were set up using primers pBAD Forward primer, pBAD1, with sequence 5'-CACACTTTGCTATGCCATAG-3' (Invitrogen Technologies, USA) and Muk-25 and run under the same conditions as has already been described in Section 3.2.1.2.

Negative control reactions without template DNA were always included. The PCR products were analyzed by resolution on 0.8% (w/v) agarose gel in order to identify pools giving an amplification product of the expected size, that is, approximately 700 bp.

3.2.2.10 Culturing of Bacteria and Small-scale Plasmid DNA Isolation

Upon identifying pools giving an amplification product of the expected size, individual colonies were aseptically picked up from the gridded plate and inoculated into 5 ml aliquots of LB broth containing 100 µg/ml carbenicillin. The bacteria were grown overnight at 37°C with shaking at 250 rpm.

Plasmid DNA was isolated using the alkaline lysis method with slight modifications as is outlined below. Four millilitres (2 x 2 ml) of each culture were centrifuged in a microfuge at 13 000 rpm for 1 min and the supernatant discarded. The bacterial pellet was re-suspended in 300 µl of buffer PI (50 mM Tris-HCl, 10 mM EDTA, pH 8.0 supplemented with 100 µg/ml RNase) by vortexing. To lyse the cells, 300 µl of a freshly prepared 200 mM NaOH, 1% SDS solution were added, mixed gently and incubated at room temperature for 5 min.

Chromosomal DNA was then precipitated by adding 300 µl of 2.55 M potassium acetate, pH 4.8 and quickly mixing by inverting the tube several times. After centrifuging at 13 000 rpm for 10 min, the supernatant (approximately 900 µl) was transferred to a new Eppendorf tube. Plasmid DNA was precipitated by adding 0.8 volumes (approximately 720 µl) of isopropanol and incubating at room temperature for 15 min. After centrifuging at 13 000 rpm for 10 min, the supernatant was discarded and the DNA pellet washed twice with 500 µl of 70% ethanol. The DNA was dried at 37°C in a dry heat block for 30 min, dissolved in 50 µl of water and then stored at -20°C.

3.2.2.11 Identification of pBADMycHisA-Wild Type Mukwa Seed Lectin Recombinants

Recombinant plasmids were identified by double digestion with *EcoRI* and *EcoRV* followed by resolution on 0.8% (w/v) agarose gel.

Five microlitres of plasmid DNA (not quantitated), 2 µl of high salt buffer, 12 µl of water, and 0.5 µl (5 units) of each enzyme were mixed together and incubated at 37°C for 2 hr. Routinely, reactions were prepared as master mixes of water, buffer and restriction endonuclease(s) and then added to the sample DNA.

EcoRV has one recognition site in the pBADMycHisA vector and none in the insert. Two DNA fragments, approximately 3.3 kb and 1.5 kb in size were therefore expected to be resolved on agarose gel.

3.2.2.12 Isolation of Plasmid DNA Using the QIAprep Spin Miniprep Kit

Plasmid DNA isolated using the alkaline lysis method is usually not pure enough for sequencing. Plasmid DNA to be sequenced was therefore isolated using the QIAprep spin miniprep kit (Qiagen, 2000a).

Colonies harbouring putative recombinants were aseptically picked up and grown overnight in 5 ml LB broth containing 100 µg/ml carbenicillin. Four millilitres (2 x 2 ml) of each culture were centrifuged in a microfuge at 13 000 rpm for 1 min and the supernatant removed. To the bacterial pellet 250 µl of Cell Suspension buffer PI supplemented with 100 µg/ml RNase were added and the cells re-suspended by vortexing. The cells were lysed by adding 250 µl of Lysis buffer P2 and mixing gently by inverting the tube five times. To precipitate chromosomal DNA, 350 µl of Neutralization buffer N3 were added and the contents of the tube mixed by inverting the tube five times. After centrifuging at 13 000 rpm for 10 min, the supernatant was applied onto a QIAprep spin column seated on a 2 ml tube. The DNA was allowed to bind onto the column by centrifuging at 13 000 rpm for 1 min. The flow-through was discarded and 500 µl of Binding buffer PB applied onto the column. After centrifuging at 13 000 rpm for 1 min, the flow-through was discarded.

To wash the column, 750 µl of Wash buffer PE were applied and the column centrifuged at 13 000 for 1 min. The flow-through was discarded and residual Wash buffer PE buffer removed by centrifuging again at 13 000 rpm for 2 min. The QIAprep column was then placed on a new Eppendorf tube and 50 µl of Elution buffer EB, pre-warmed to 65°C, applied onto the column. After incubating at room temperature for 10 min, the DNA was eluted by centrifuging at 13 000 rpm for 1 min.

Five microlitres of the plasmid DNA preparation were digested with the appropriate restriction endonuclease(s) and the DNA resolved on 0.8% (w/v) agarose gel to check for quality and to estimate quantity.

3.2.3 Sequencing of Cloned Wild Type Mukwa Seed Lectin DNA Fragments

The cloned fragments were sequenced by the Sanger method (1977) using the thermo-sequenase radiolabelled terminator cycle sequencing kit (United States Biochemical Corporation, 2000). Sanger's technique uses controlled synthesis of DNA to generate

fragments that terminate at specific points along the target sequence. Thermo-sequenase efficiently incorporates the radiolabelled dideoxynucleotides and hence very low amounts of isotope are used. Moreover, the use of this thermostable DNA polymerase allows for cycle sequencing, the process of using repeated cycles of thermal denaturation, primer annealing and polymerization to produce large amounts of product in a DNA sequencing reaction.

The pBAD*MycHisA*-wild type mukwa seed lectin recombinant plasmid DNA isolated using the QIAprep spin miniprep isolation kit (Qiagen, Germany) was used for sequencing. Each of the following primers: pBAD Forward primer (pBAD1), Muk-6 (sequence 5'-CATCGCACCGCCGGATACTAC-3') and Muk-25 were used respectively in order to sequence the cloned fragments entirely.

3.2.3.1 Preparation of Sequencing Reactions

Sequencing reactions were prepared in three stages. Firstly, the chain extension reaction mixture and the termination mixture were prepared separately. After adding the reaction mixture to each nucleotide termination reaction tube, the mixture was then subjected to PCR. The substitution of the nucleotide analog dITP for dGTP was meant to help eliminate compressions due to secondary structures as dITP forms weaker base pairs with dCTP that are more readily denatured before separation by gel electrophoresis.

3.2.3.1.1 Reaction Mixture

Approximately 0.5 µg of plasmid DNA is required for each sequencing reaction. To the plasmid DNA, 2 µl of 10 x reaction buffer, 1 µl of a 2.5 µM solution of primer, 2 µl (8 units) of thermo-sequenase and water up to a total volume of 20 µl were added and mixed well. The 10 x reaction buffer consisted of 260 mM Tris-HCl, pH 9.5, and 65 mM MgCl₂. Four and a half microlitres of the reaction mixture were then aliquoted into each of four different

coloured 0.5 ml PCR tubes. The yellow tube represented the G, colourless, the A, red, the T and blue the C reaction respectively.

3.2.3.1.2 Termination Mixture

Four different Eppendorf tubes were labelled G, A, T and C respectively. Two and a half microlitres of the dITP nucleotide master mix (7.5 μ M dATP, 7.5 μ M dCTP, 7.5 μ M dTTP and 13.5 μ M dITP) plus 0.5 μ l of one of the four [α - 33 P]-dideoxy analogues were required for each reaction tube. Master mixes of the termination mixture were therefore routinely prepared. The appropriate amount of the dITP nucleotide master mix was added to each of the labelled four tubes and mixed with the appropriate dideoxy analogue. For example, for 10 reactions, 25 μ l of the dITP nucleotide master mix were added to each of the G, A, T and C tubes, followed by 5 μ l of each of [α - 33 P]-ddGTP, [α - 33 P]-ddATP, [α - 33 P]-ddTTP and [α - 33 P]-ddCTP respectively. After mixing well, 2.5 μ l of the appropriate termination mixture were then added to the 4.5 μ l of the reaction mixture and centrifuged at maximum speed for 30 sec to bring the contents to the bottom of the tube.

3.2.3.1.3 PCR

After mixing the reaction and termination mixes, the PCR was carried out using an automated PCR protocol. Twenty cycles of template denaturation at 94°C for 15 sec, primer annealing at 55°C for 30 sec and primer extension at 60°C for 10 min were run in a programmable thermocycler. The reaction was stopped by adding 4 μ l of the stop solution (loading buffer) to each tube. The stop solution consisted of 95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol FF. The reactions were centrifuged at maximum speed for 30 sec to bring the contents to the bottom of the tube, denatured by heating at 80°C for 10 min, snap-cooled on ice and immediately loaded onto a gel or were frozen at -20°C for use within a week.

3.2.3.2 Denaturing Polyacrylamide Gel Electrophoresis

The nested series of DNA fragments generated in the sequencing reactions was resolved by high voltage electrophoresis (Consort Model 3 000 x i programmable power supply, Consort, Belgium) on 5% (w/v) denaturing polyacrylamide gels.

3.2.3.2.1 Gel Casting

Gel sandwiches were prepared using a full size plate and one with round edges. Both plates were thoroughly washed with laboratory detergent and tap water and then rinsed with distilled water. The full size plate was treated by scrubbing it with the following solutions one after the other: saturated KOH in 95% technical alcohol; 95% technical alcohol; technical acetone; Wacker solution (1 ml Wacker, 190 ml pure ethanol, 4 ml water and 6 ml 10% acetic acid); 95% technical alcohol; Wacker solution and finally 95% technical alcohol. Wacker solution causes the gel to stick onto this plate. The plate with round edges was treated by scrubbing it with the following solutions one after the other: saturated KOH in 95% technical alcohol; 95% technical alcohol; technical acetone and twice with dimethyldichlorosilane. The toxic dimethyldichlorosilane was used to siliconize the glass plate so that the gel would not stick onto this plate.

The 20 cm x 50 cm glass plate sandwich was made using 0.4 mm spacers. Bulldog clamps were used to keep the spacers in place. The gel was prepared by dissolving 50 g of urea, 4.8 g acrylamide, 0.2 g N,N'-methylene-bis-acrylamide, 0.6 g boric acid and 1.2 g trizma base in 50 ml of water and then the solution made up to 100 ml with water. Just before pouring the gel, 200 µl of 10% ammonium persulphate and 100 µl of N,N,N',N'-tetramethylethylenediamine (TEMED) were added and the gel quickly poured into the gel sandwich laid on a flat surface. A 50-well comb was inserted at the top of the gel and held in place using clamps. Polymerization took place within 30 min but was routinely left to proceed overnight.

3.2.3.2.2 Electrophoresis

After polymerization, the comb was removed and the top of the gel rinsed with 0.5 x TB buffer (89 mM Tris base and 89 mM boric acid, pH 8.0). To ensure even conduction of heat generated during electrophoresis, an aluminium plate (0.4 cm thick, 34 x 22 cm) was clamped onto the front glass plate with the same clamps used to hold the gel sandwich together. The gel assembly was then set up and placed in the lower buffer chamber. The lower buffer was prepared by mixing 200 ml TB buffer and 100 ml of 3 M sodium acetate, pH 5.2, and poured into the lower buffer chamber. The upper buffer chamber was filled with 0.5 x TB buffer. The gel was pre-run at 25 mA for 30 min.

The wells were rinsed with TB buffer to remove dissolved urea and 3 µl of each of the denatured sequencing reaction loaded into each well in the order GATC. The gel was run at 2 000 V, 25 mA and 150 W for 5 hr.

3.2.3.2.3 Gel Fixing, Drying and Autoradiography

The gel, still on the plate treated with Wacker solution, was fixed in 5 L of 10% acetic acid for 30 min to remove the urea. After fixing, the gel was rinsed three times with distilled water and left to air-dry overnight. The dry gel was exposed to Fuji Medical X-ray 100 NIF film overnight and developed in a dark room in a developing machine (AGFA, Germany) using AGFA reagents prepared according to the manufacturer's instructions.

3.2.3.3 Reading of Sequences

Sequencing autoradiographs were read and entered manually using the PC/GENE Sequence Analysis software.

3.2.4 Expression of the Mature Wild Type Mukwa Seed Lectin in *E. coli*

The expression vector pBADMycHisA requires host cells that are capable of transporting L-arabinose but incapable of metabolizing it. These conditions are important as a constant level of arabinose, the inducer of expression, can then be maintained in the host cells.

3.2.4.1 Transformation of *E. coli* TOP10 Cells Using the CaCl₂ Method

Two separate transformation experiments were set up. The first experiment was a single transformation of *E. coli* TOP10 cells with the pBADMycHisA-wild type mukwa seed lectin clone. One microlitre (approximately 100 ng) of the plasmid DNA was added to 200 µl of thawed *E. coli* TOP10 competent cells and mixed gently. The mixture of plasmid DNA and host cells was incubated on ice for 10 min followed by incubation at 37°C for 5 min. One millilitre of LB broth was then added to the mixture and 100 µl of the cell suspension plated out on LB agar supplemented with 100 µg/ml of carbenicillin. The plate was incubated at 37°C overnight.

In the second experiment, double transformation of *E. coli* TOP10 cells with the pBADMycHisA-wild type mukwa seed lectin clone as well as BL21-CodonPlusRIL was carried out. BL21-CodonPlusRIL enhances expression of codons that are rarely used in *E. coli*. One microlitre (approximately 100 ng) of each of the two plasmid DNA preparations was added to 200 µl of thawed *E. coli* TOP10 competent cells and mixed gently. The plasmid DNA-host cells mixture was incubated on ice for 10 min followed by incubation at 37°C for 5 min. One millilitre of LB broth was then added to the mixture and 100 µl of the cell suspension plated out on LB agar supplemented with both 100 µg/ml of carbenicillin and 25 µg/ml chloramphenicol. The plate was incubated at 37°C overnight.

Individual colonies were selected from each of the single transformants and double transformants plates and streaked on fresh LB agar plates supplemented with the appropriate antibiotic(s) in order to obtain pure colonies. The plates were incubated at 37°C overnight.

3.2.4.2 Pilot Expression of the Wild Type Mukwa Seed Lectin

Each recombinant protein has different characteristics that may affect optimum expression. The optimal level of expression can be determined by varying the amount of L-arabinose over a ten thousand-fold range or by performing a time course.

For each transformant, 2 ml of LB broth containing the appropriate antibiotic(s) were inoculated with one recombinant *E. coli* colony. In other words, for single transformants the LB broth contained 100 µg/ml carbenicillin and for double transformants the LB broth contained both 100 µg/ml carbenicillin and 25 µg/ml chloramphenicol. The cultures were grown with shaking at 250 rpm overnight at 37°C.

The next day, for each culture, five tubes were labelled 1 through 5 and 10 ml of LB broth containing the appropriate antibiotic(s) added to each tube. Each tube was inoculated with 100 µl of the overnight culture and the cultures grown at 37°C with shaking to an $OD_{660nm} = 0.40 - 0.45$. While the cells were growing, four 10-fold serial dilutions of 20% (w/v) L-arabinose were prepared aseptically. A 1 ml aliquot of cells in their mid-log phase was removed from each tube, centrifuged at 13 000 rpm in a microfuge for 1 min and the supernatant aspirated. The cell pellets were frozen at -20°C and these were the zero time point (ztp) or control samples.

One hundred microlitres of each of a 0.002%, 0.02%, 0.2%, 2% and 20% L-arabinose solution were added to tubes 1 through 5 respectively. The final concentration of L-arabinose in tubes 1 through 5 was therefore 0.00002%, 0.0002%, 0.002%, 0.02% and 0.2% respectively. The cultures were grown at 37°C with shaking at 250 rpm and 1 ml aliquots taken out of each tube after 4 hr and after 16 hr respectively. The 1 ml aliquots were centrifuged, the supernatant aspirated and the pellets frozen as before.

3.2.4.3 Analysis of Expression of the Wild Type Mukwa Seed Lectin by SDS-PAGE

When proteins are boiled in the presence of the anionic detergent sodium dodecyl (lauryl) sulphate (SDS), all non-covalent bonds are disrupted resulting in proteins with an overall negative charge and a helical rod-like conformation. Moreover, if a reducing agent such as β -mercaptoethanol or dithiothreitol is included in the protein sample, all inter- and intra-disulphide bonds are broken giving the protein in its subunits. As the proteins consequently have the same charge and shape, separation by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is on the basis of size only.

If total protein in *E. coli* expressing the lectin of interest is resolved by SDS-PAGE and the gel stained with Coomassie blue, the band pattern can be visualized. The Tris-Tricine-SDS-PAGE (Schägger and Von Jagow, 1987) method for the separation of proteins in the range from 1 to 100 kDa was used to analyze the expression of the wild type mukwa seed lectin.

3.2.4.3.1 Gel Casting

Gels were poured and run in a Bio-Rad Mini-Protean II Dual Slab cell (BIO-RAD, USA). For two gels, a 10% resolving (running) gel was prepared by mixing 1.62 ml of solution 1A (60% {w/v} acrylamide + 1.88% {w/v} N,N'-methylene-bis-acrylamide), 3.3 ml gel buffer (3 M Tris-HCl, pH 8.45), 1.09 ml glycerol, 3.86 ml double distilled water, 100 μ l of 10% (w/v) SDS, 50 μ l of 10% (w/v) ammonium persulphate and 5 μ l of N,N,N',N'-tetramethylethylenediamine (TEMED). The resolving gel solution was then poured into the plate assembly and immediately overlaid with water to exclude oxygen. The unused gel solution was saved for 10 to 20 min to check when polymerization had occurred.

When the gel had set, the overlaying water was removed. The 4.5% stacking gel was prepared by mixing 0.4 ml of solution 1A, 1.24 ml of gel buffer, 3.32 ml of double distilled water, 40 μ l of 10% (w/v) SDS, 25 μ l of 10% (w/v) ammonium persulphate and 2.5 μ l of

TEMED. Immediately after pouring the stacking gel into the plate assembly, well-forming combs were placed at an angle to prevent trapping air bubbles under the teeth of the combs.

3.2.4.3.2 Sample Preparation and Electrophoresis

Each cell pellet collected during the pilot expression experiment was re-suspended in 100 μ l of 1 x SDS-PAGE sample buffer (Laemmli, 1970). The 1 x SDS-PAGE sample buffer consisted of 4% (w/v) SDS, 8.3% (v/v) glycerol, 50 mM Tris-HCl, pH 6.8, 10% (v/v) β -mercaptoethanol and 0.01% (w/v) Coomassie blue G-250. The cell suspensions and the protein molecular mass standards were placed in boiling water for 5 min and then centrifuged briefly.

When the stacking gel had set, the combs were removed and the plates fitted into the central stand of the running apparatus. The upper chamber was filled with cathode buffer (0.1 M Tris, 0.1 M Tricine and 0.1% {w/v} SDS) and checked to ensure that there was no leakage occurring. The stand was placed into the tank and anode buffer (0.2 M Tris-HCl, pH 8.9) added to the lower chamber to cover the lower electrode.

Five microlitres of each sample and the standard were loaded onto the gel, the apparatus connected to the power pack and electrophoresis started at 30 volts (25 mA). After 45 min, the voltage was increased to 80 volts (60 mA), and 15 min later it was increased to 90 volts and kept constant until the tracking dye reached the end of the gel.

After electrophoresis, the gel was stained in Coomassie blue stain (0.1% {w/v} Coomassie brilliant blue R-250, in 50% {v/v} methanol and 10% {v/v} acetic acid) for 1 hr with gentle shaking at room temperature. The gel was then destained in destaining solution (25% {v/v} methanol and 10% {v/v} acetic acid), rinsed in distilled water and the band pattern observed.

3.2.4.4 Analysis of Expression of the Wild Type Mukwa Seed Lectin by Western Blotting

After SDS-PAGE, the protein bands can be transferred from the gel onto the surface of a thin matrix such as nitrocellulose or nylon membrane. As the proteins are immobilized on the matrix and are readily accessible to interactions with ligands, if protein-specific polyclonal or monoclonal antibodies are available, detection by immunoprobings is possible. The protein molecular mass standards used in this technique must, however, be pre-stained or biotinylated. The biotinylated standards are not boiled but are first incubated at 37°C for 2 hr.

Western blotting was used to verify if anti-native mukwa seed lectin antibodies could recognize the protein that had been expressed in *E. coli* TOP10 cells. The electrophoretic transfer of the total protein in the *E. coli* TOP10 cells was done using Towbin and co-workers' method (1979) in a Transblot apparatus (Bio-Rad, UK).

3.2.4.4.1 Electrophoretic Transfer

After SDS-PAGE, the stacking gel was removed, the orientation noted by cutting a corner and the gel equilibrated by gently shaking in transfer buffer for 15 min at room temperature. The transfer buffer consisted of 25 mM Tris, 192 mM glycine and 20% (v/v) methanol, pH 8.3. Washing in transfer buffer removes SDS and other gel constituents that may interfere with transfer.

The gel sandwich was assembled in a glass tank filled with sufficient transfer buffer to immerse it. The grey panel was placed horizontally in the glass tank and a pre-soaked fiber pad placed on top. Two sheets of filter paper (Whatman, 3 MM), pre-cut to the same size as the gel and saturated with transfer buffer, were placed on top of the fiber pad and the gel gently laid on top of the sheets of filter paper. The uncovered side of the gel was overlaid with a pre-wetted nitrocellulose sheet (Schleicher and Schuell, NC-0.45 µm), pre-cut to the size of the gel plus 1 to 2 mm on each edge. Air bubbles were removed by rolling over a clean test tube or glass rod. The nitrocellulose was overlaid with another two sheets of buffer-saturated filter

paper and another pre-soaked fiber pad placed on top. The sandwich was closed and immediately immersed in transfer buffer in the electroblotting apparatus in the correct orientation. The grey panel was placed towards the cathode of the apparatus such that the nitrocellulose was positioned on the anode side of the gel. Transfer was achieved by applying a voltage of either 30 volts (approximately 75 mA) overnight or 100 volts (approximately 300 mA) for 2 hr with an ice-cold circulating water bath used to cool the system.

After electrophoresis, the nitrocellulose sheet was removed from the blotting apparatus and the orientation noted by marking with a soft lead pencil.

3.2.4.4.2 Immunoprobng and Visual Detection of Proteins

The Western blot was placed in a plastic tray and non-specific ligand binding sites blocked by shaking in 3% (w/v) bovine serum albumin (BSA) dissolved in phosphate buffered saline (PBS), pH 7.3 for 1 hr on a rocking platform. The PBS consisted of 10 mM potassium phosphate, pH 7.3 containing 0.15 M NaCl.

After blocking, the blot was washed twice in 0.05% (v/v) Tween-20 in PBS. Each washing step involved placing the blot in the appropriate solution and gently shaking for 10 min. After washing, the blot was incubated at room temperature, overnight, in a 1: 500 dilution of the rabbit anti-serum directed against the mukwa seed lectin. The antibody solution was prepared in 3% (w/v) BSA and 0.01% (v/v) Tween-20 dissolved in PBS. The blot was washed five times in 0.05% (v/v) Tween-20 in PBS and then incubated for 1 hr in a solution containing 1 µg/ml protein A-biotin or anti-rabbit IgG alkaline phosphatase conjugate and 2 µg/ml extravidin-phosphatase in PBS. The blot was washed five times in 0.05% (v/v) Tween-20 in PBS and twice (in 5-min incubations) in PBS only.

The blot was developed with 5-bromo-4-chloro-3-indoyl phosphate, *p*-toluidine salt (BCIP) and nitroblue tetrazolium (NBT) for chromogenic visualization of the bands. This stain was prepared just before use by adding 330 µl NBT and 165 µl BCIP to 50 ml of alkaline phosphate buffer (100 mM Tris-HCl, pH 9.5, 10 mM MgCl and 100 mM NaCl). The NBT and

BCIP solutions were prepared in separate glass tubes by dissolving 50 mg in 1 ml of 70% (v/v) and 1 ml of 100% (v/v) of dimethylformamide (DMF) respectively. Bands usually appeared after 5 to 15 min in the staining solution and washing the blot in distilled water terminated the reaction. The blot was then air-dried and kept in aluminium foil.

3.2.5 Assaying the Expressed Wild Type Mukwa Seed Lectin for Activity

The binding capacity of the expressed lectin was initially tested using the crude cell extract. After successful expression, the experiment was then scaled-up to obtain large quantities of the protein.

3.2.5.1 Scale-Up of Expression for Assaying for Biological Activity

The conditions determined in the pilot expression experiment were used to grow and induce 50 ml of recombinant *E. coli* TOP10 cells. Two millilitres of LB broth containing 100 µg/ml carbenicillin were inoculated with one expressing recombinant *E. coli* colony and the culture grown at 37°C overnight, with shaking at 250 rpm.

The next day, 50 ml of LB broth, containing 100 µg/ml carbenicillin, were inoculated with 1 ml of the overnight culture and the culture grown at 37°C, with shaking at 250 rpm, to an OD_{660nm} = 0.40-0.45. The optimal amount of L-arabinose (500 µl of a 20% solution) was added to the culture to induce expression. The culture was grown at 37°C, with shaking at 250 rpm, until the optimal time point was reached (16 hr).

The cells were harvested by centrifuging at 3 000 x g for 10 min at 4°C (Sorvall RC-5B Refrigerated Superspeed Centrifuge, Du Pont Instruments, USA). The cell pellet was washed in 10 ml of saline azide solution (0.15 M NaCl + 0.02% {w/v} NaN₃) supplemented with 5 mM CaCl₂ and 5 mM MnCl₂, and then re-suspended in 2 ml of the same solution. The Ca²⁺ and Mn²⁺ ions are required for the activity of the mukwa seed lectin.

3.2.5.2 Sonication of Cells

Agitating bacterial cells by high-frequency sound waves breaks the outer cell walls thereby releasing their contents. The cells, in saline azide supplemented with 5 mM CaCl₂ and 5 mM MnCl₂, were sonicated using amplitude of approximately 14 milliamplicons (Soniprep 150, MSE Scientific Instruments, UK). As the technique generates heat, samples were kept on ice during sonication and this was done for 6 x ½ min with ½ min intervals.

After sonication, the samples were centrifuged in a microfuge at 13 000 rpm for 10 min and the supernatant kept at 4°C.

3.2.5.3 Biological Activity Tests

Lectins are capable of agglutinating cells and the agglutinating activity of the lectin is often inhibited by simple monosaccharides. Red blood cells have a range of sugar moieties on their surface and are preferentially used for such tests. The biological activity of the crude cell extract was therefore assessed by haemagglutinating activity tests followed by inhibition of agglutination by methyl- α -D-mannopyranoside using rabbit and untreated human erythrocytes.

3.2.5.3.1 Washing of Rabbit and Human Erythrocytes

Rabbit blood was collected and mixed with an equal volume of Alsever solution (0.05 g glucose, 0.80 g Na-citrate, and 0.42 g NaCl in 100 ml bidistilled water) containing 2 drops of heparin per 5 ml of the solution. Human blood was collected in heparin-coated tubes.

To wash the erythrocytes, 2 ml of rabbit blood (or 1 ml of human blood) were placed in a 30-ml centrifuge tube, saline azide solution added to the mark, and the cells collected by centrifugation at 450 x g (1 800 rpm in the SS34 rotor) for 20 min at 4°C (in the cold room). Washing was repeated twice more using fresh saline azide solution each time. After the third wash, the erythrocyte pellet was re-suspended in 25 ml saline azide solution to give a 4% (v/v) suspension.

3.2.5.3.2 Haemagglutination Assays

Haemagglutinating activity was measured in U-shaped microtitre plates. The wells of the first column on the plate contained 25 μ l of saline azide solution only and therefore acted as the negative blank. The wells of the second column contained 25 μ l of the undiluted lectin sample. Twenty-five microlitres of the undiluted lectin sample were added to the wells of the third column and the lectin sample serially diluted with an equal volume of saline azide solution to give a final volume of 25 μ l of diluted lectin sample per well. Lastly, 50 μ l of the 4% (v/v) erythrocyte suspension were added to all the wells on the plate except for the last column. This meant that if necessary, further dilutions could be started from this column. The plates were then incubated at room temperature for 1 hr and the end-point of the titration (agglutination titre) estimated visually and expressed as the highest dilution that gave visible agglutination.

3.2.5.3.3 Inhibition of Agglutination

The mukwa seed lectin is mannose-specific and hence methyl- α -D-mannopyranoside, a more effective inhibitor than mannose for mannose/glucose-specific legume lectins (Nomura *et al.*, 1998), was used in the inhibition of agglutination test.

Firstly, the haemagglutinating activity of the crude cell extract was determined and the lectin sample diluted to give exactly 4 agglutinating units, that is, 3 wells in which agglutination was visible. This diluted lectin sample (crude cell extract) was then used throughout to test for the inhibition of agglutination.

The wells of the first column of the U-shaped plate contained 25 μ l of saline azide solution only (no lectin and no sugar) and therefore acted as the negative blank. The wells of the second column contained 25 μ l of the lectin sample only but no sugar and thus acted as the positive blank. The wells of the third column contained 25 μ l of the undiluted sugar (200 mM in saline azide). From the fourth well of each row, 25 μ l of the sugar were serially diluted with

equal volumes of saline azide solution to give a final volume of 25 μ l of the diluted sugar. Twenty-five microlitres of the lectin sample were then added to each well except for the blanks. Plates were incubated at room temperature for 20 min to allow the sugar to bind to the lectin. Lastly, 50 μ l of the 4% (v/v) erythrocyte suspension were added to all the wells on the plate except for the last column (to allow for further dilutions if necessary).

Plates were incubated at room temperature for 1 hr, at the end of which inhibition of agglutination was estimated visually and expressed as the lowest concentration of sugar able to inhibit agglutination.

3.2.6 Purification of the Expressed Wild Type Mukwa Seed Lectin by Affinity Chromatography

In order to analyze and characterize a given protein, reasonably large quantities of the pure protein must be isolated.

3.2.6.1 Scale-Up of Expression for Purification

Optimal conditions determined in the pilot experiment (Section 3.2.4.2) were used to induce 2 x 500 ml of cells. As foreign proteins produced in *E. coli* are at times rapidly degraded by bacterial proteases (Gottesman, 1996), the two cultures were grown separately so that the cell extracts could be processed either in the presence or in the absence of a protease inhibitor cocktail respectively.

Thirty millilitres of LB broth containing 100 μ g/ml carbenicillin were inoculated with one expressing recombinant *E. coli* TOP10 colony and grown overnight at 37°C, with shaking at 250 rpm. The next day 10 ml of the overnight culture were used to inoculate each 500 ml of LB broth supplemented with the appropriate antibiotic. When the cells had grown to an $OD_{660nm} = 0.40-0.45$, 5 ml of a 20% solution of L-arabinose were added to each 500 ml culture to induce expression and the cultures allowed to grow overnight.

The next day, one millilitre was taken from each of the two cultures, centrifuged at 13 000 rpm for 1 min and the pellet kept at -20°C for SDS-PAGE analysis. The rest of the cultures were centrifuged at 3 000 x g (4 200 rpm in Sorvall GSA rotor) for 20 min.

The pellet from one culture was washed in 50 ml of saline azide solution supplemented with 5 mM Mn^{2+} , 5 mM Ca^{2+} and a freshly prepared protease inhibitor cocktail (1 mM phenylmethylsulfonylfluoride {PMSF}+ 2 μ M leupeptin + 2 μ M pepstatin A). The cells were re-suspended in 20 ml of the solution used for washing.

The pellet from the second culture was washed in 50 ml of saline azide solution supplemented with 5 mM each of Mn^{2+} and Ca^{2+} ions only. The cells were re-suspended in 20 ml of the same solution.

The cells were sonicated 6 x ½ min with ½ min intervals as has already been described in Section 3.2.5.2. Cell walls were removed by centrifuging for 10 min at 15 000 rpm (Sorvall SS34 rotor). The supernatant was applied onto a mannose-Sepharose-4B column equilibrated in saline azide solution.

3.2.6.2 Affinity Chromatography

Lectins are routinely purified by affinity chromatography. The technique is based on the fact that a ligand (saccharide), covalently immobilized onto a solid support, is only recognized by a limited number of proteins in the extract. Adding free ligand into the elution buffer can then specifically elute the adsorbed protein(s). Since the mukwa seed lectin is mannose-specific, the lectin can be isolated by affinity chromatography on mannose-Sepharose-4B.

3.2.6.2.1 Immobilization of Mannose onto Sepharose-4B

Sugar molecules can be coupled to Sepharose resins after activation of the beads with divinylsulfone. Divinylsulfone activation gives rise to a very stable linkage between ligand and matrix. Mannose was immobilized onto Sepharose-4B according to Porath (1974).

Twenty grams of Sepharose-4B (wet weight) were washed on a sintered glass funnel with double distilled water followed by 100 ml of 1 M Na_2CO_3 buffer, pH 11. The beads were re-suspended in 20 ml of the latter buffer.

To activate the Sepharose-4B, 4 ml of divinylsulfone were added to the suspension and the gel shaken for exactly 90 min at 25°C. The gel was transferred onto a sintered glass funnel and washed extensively with 1 L of double distilled water followed by 0.5 L of 1 M Na_2CO_3 buffer, pH 10. The washed beads were added to 20 ml of a 100 mg/ml D-mannose solution, prepared by dissolving 2 g of the saccharide in 1 M Na_2CO_3 buffer, pH 10. The gel suspension was allowed to shake for 2 hr at 50°C and coupling left to proceed by shaking overnight at 35°C.

The gel suspension was transferred onto a sintered glass funnel and washed consecutively with 0.5 L each of: 1 M Na_2CO_3 buffer containing 1 M NaCl, pH 10; 0.2 M glycine-HCl buffer containing 1 M NaCl, pH 3; 1 M NaCl and finally with 1 L of double distilled water. To block the remaining reactive groups the gel was re-suspended in 100 ml of 0.2 M Tris-HCl buffer, pH 8.5, and washed with shaking at 25°C for 3 hr. The gel was transferred to a sintered glass funnel and washed, first with 1 L of double distilled water and then with 1 L of 0.1 M NaHCO_3 buffer, pH 8.2, and stored in saline azide solution.

3.2.6.2.2 Affinity Chromatography on Mannose-Sepharose-4B Column

The mannose-Sepharose-4B column (2 x 7.5 cm) was initially equilibrated with saline azide solution at a flow rate of 20 ml/hr using a pump with the absorbance of the eluent being monitored at 280 nm. The crude cell extract (18.5 ml) was loaded onto the column and the

unbound protein washed with saline azide solution until the absorbance at 280 nm approached the set base line. The bound lectin was eluted with 0.3 M mannose in saline azide and the remaining protein with 0.1 M acetic acid. The peaks (eluted protein) were separately dialyzed three times in 2 L of saline azide overnight at 4°C. The column was regenerated by washing with saline azide and stored at room temperature in the same solution.

3.2.6.3 Analysis of Purification

The purification process was analyzed by testing the crude cell extract, the unbound fraction and the bound fraction(s) for agglutinating activity and running tricine-SDS-PAGE as has already been described in Section 3.2.4.3. All the fractions were first dialyzed against water to remove saline azide and the α -mannose, and the bound fraction was then concentrated using a Millipore Ultrafree-15 Centrifugal Filter device of 10 000 MW limit (Sigma, UK) before loading onto the gel. Protein concentrations in each fraction were determined by the bicinchoninic acid (BCA) protein assay according to Smith *et al.*, (1985) using bovine serum albumin as a standard. The estimated protein concentrations and the agglutinating activity were then used to calculate the total activity and the specific activity of the lectin.

3.2.6.3.1 Determination of Protein Concentration by the BCA Assay

The bicinchoninic acid (BCA) protein assay is based on chemical principles similar to those of the Biuret and Lowry assays. The protein mixture to be analyzed is reacted with Cu^{2+} to give Cu^+ . The Cu^+ is chelated by BCA, which converts the greenish colour of the BCA to the purple colour of the Cu^+ -BCA complex whose absorbance can be read at 562 nm.

The BCA Protein Assay Reagent kit from Pierce (Rockford III, USA) was used to estimate the protein concentration in each lectin sample collected throughout the purification procedure. The assay was performed in flat-bottomed microtitre plates. All samples and standards were diluted in saline azide solution. A series of bovine serum albumin (BSA)

solutions of known concentrations ranging from 0.2 to 2.0 mg/ml was prepared from a 2 mg/ml stock solution and used as standards.

Two hundred microlitres of the BCA working reagent (prepared by mixing 50 parts reagent A containing sodium carbonate, sodium bicarbonate, BCA detection reagent and sodium tartrate in 0.2 N NaOH with 1 part reagent B containing 4% copper sulphate solution) were added to 10 μ l of each protein sample. A negative blank, containing 10 μ l of saline azide solution and 200 μ l of BCA working reagent, was included on the plate. The plate was gently shaken to mix the contents in each well, the plate covered and incubated at 37°C for 30 min. After incubation, absorbance was read at 595 nm on a Bio-Rad/ 3550 microtitre reader. The protein concentration of the lectin samples was then estimated from the BSA standard curve.

3.2.6.4 Determination of the Effect of Culture Volumes on the Yield of Expressed Lectin

During scaling-up, the amount of expressed protein obtained from a larger volume of culture can be less than expected relative to the amount determined by the pilot expression experiment. The effect of culture volumes on the amount of the expressed lectin was therefore investigated.

Optimal conditions determined in the pilot experiment (Section 3.2.4.2) were used to grow and induce 10 x 50 ml of cells and 1 x 500 ml of cells with 20% L-arabinose as described in Sections 3.2.5.1 and 3.2.6.1 respectively. During harvesting, the 10 x 50 ml cultures were pooled together and treated as one sample while the cells from the 1 x 500 ml culture were treated as a second sample. The cells from each sample were re-suspended in saline azide supplemented with 5 mM Ca^{2+} and 5 mM Mn^{2+} ions, 1/25 volume of the original culture, sonicated, centrifuged and the supernatant passed through a mannose-Sepharose-4B column.

3.2.7 Mutagenesis of the Sugar-Binding Specificity Loop of the Mukwa Seed Lectin

Introducing specific mutations into the DNA coding for a given protein will enable one to follow their effect on the functional properties of the protein. Mutagenesis experiments were carried out in order to establish if changing the amino acid sequence of the mukwa seed lectin specificity loop would also change the specific sugar that this lectin binds. Clone Muk151QII28, pUC18 carrying the cDNA encoding the wild type mukwa seed lectin, was available as the template for the *in vitro* mutagenesis.

3.2.7.1 Mutagenic Primer Design

All mutagenic primers designed had a restriction endonuclease recognition site incorporated into them. This meant that even though the specific amino acid sequence was maintained, in some cases the actual nucleotide sequence had to be changed in order to provide a restriction endonuclease recognition site. Identification of putative mutants would therefore be possible by restriction endonuclease analysis on agarose gel.

In order to try and change the sugar specificity of the mukwa seed lectin, two different approaches were used. The first approach involved removing the DNA segment coding for seven amino acid residues that constitute the part of the mukwa seed lectin specificity loop that interacts with the sugar. The excised segment was replaced with the respective DNA segment coding for either nine or ten amino acid residues of the corresponding regions of the *Ulex europaeus* II lectin (UEA II) or the *Erythrina corallodendron* lectin (ECorL) specificity loops (see Figure 3.3).

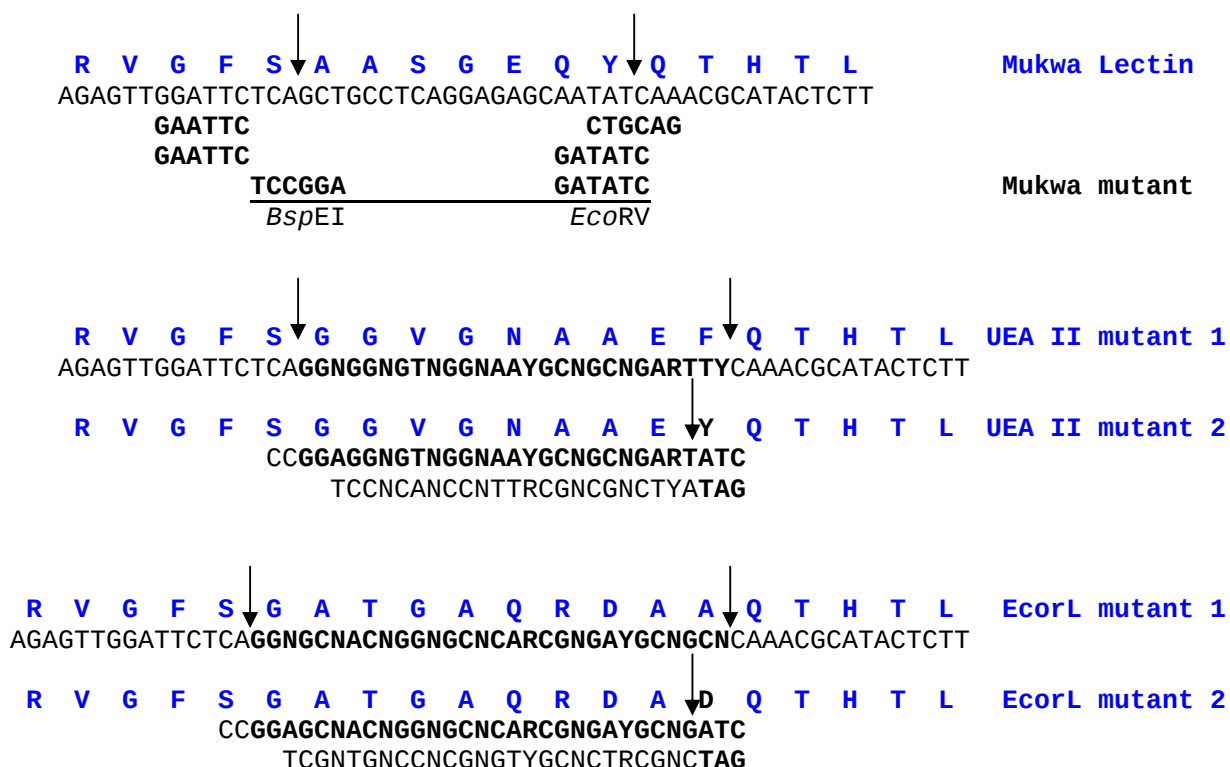


Figure 3.2: Mutagenic primer design showing the removal of the DNA segment coding for the part of loop D of the mukwa seed lectin that interacts with the sugar and its replacement with the corresponding region of either the UEA II or ECorL specificity loops.

The amino acid residues of the UEA II and ECorL specificity loops were designated USP1 through USP9 and ESP1 through ESP10 respectively. In this approach, seven mutagenic primers designated Muk-17, Muk-18, Muk-19, Muk-20, Muk-21, Muk-22 and Muk-35 were ordered from Invitrogen Life Technologies, USA. Muk-17 and Muk-18 were designed based on the clone Muk151QII28 (GenBank Accession Number AJ426056). Muk-19 and Muk-20 were designed based on the UEA II gene sequence (Loris *et al.*, 2000) while Muk-21, Muk-22 and Muk-35 were based on the ECorL gene sequence (Arango *et al.*, 1990).

Muk-17, with sequence 5'-CCTCAGGAGAGCGATATCAAACGCATA-3', has an *EcoRV* recognition site, shown in bold, while Muk-18, with sequence 5'-GGGTTAGAGTTGGATTCTCCGGAGCCTCAGGAGAGCGATATC-3' has a *BspEI* recognition site, shown in bold. These two mutagenic primers were used for the introduction of the *EcoRV* and *BspEI* recognition sites at the 3' and 5' ends respectively of the specificity

loop in clone Muk151QII28. Digestion of the resulting clone, Muk151QII28-Mut, with *EcoRV* and *BspEI* would then allow the nucleotide sequence coding for the part of the specificity loop of the mukwa seed lectin that interacts with the sugar to be removed.

Muk-19, with sequence 5'-CCGGAGGCGTTGGCAACG**CGGCCGAAT**-3', has a *BstZI* (*EagI*) recognition site, shown in bold, while Muk-20 has the complementary sequence 5'-ATTCGGCCGCGTTGCCAACGCCT-3'. Annealing these two primers would give a DNA segment specifying for the part of the UEA II specificity loop that interacts with the sugar. Introduction of this DNA segment into the *EcoRV*-*BspEI*-digested clone Muk151QII28-Mut would give the amino acid residue USP9 as tyrosine, Y (UEA II mutant 2) instead of the intended phenylalanine, F (UEA II mutant 1) [see Figure 3.2].

Muk-21, with sequence 5'-CCGGAGCG**ACCGGT**GCGCAGCGTGATGCGG-3', has an *AgeI* recognition site, shown in bold, while Muk-22 has the complementary sequence 5'-CCGCATCACGCTGCGCACCGGTGCT-3'. Annealing these two primers would give a DNA segment specifying for the part of the ECorL specificity loop that interacts with the sugar. However, introduction of this DNA segment into the *EcoRV*-*BspEI*-digested clone Muk151QII28-Mut would give the amino acid residue ESP10 of the ECorL specificity loop as aspartic acid, D (ECorL mutant 2) instead of the required alanine, A (ECorL mutant 1) [see Figure 3.2]. The nucleotide sequence coding for the D residue would then be changed to the nucleotide sequence coding for an A residue using Muk-35. Muk-35, with sequence 5'-GGAGCGACCG**GTGCAC**AGCGTGATGCGGCTCAAACGCATACTCTT-3', has an *ApaLI* recognition site, shown in bold and the targeted A codon, underlined.

The second approach involved introducing specific mutations predicted from X-ray crystallography to result in a change of the sugar specificity from α -mannose/glucose to β -mannose/glucose (Loris *et al.*, 2004). In this approach three mutagenic primers, designated Muk-30, Muk-33 and Muk-34, were designed based on clone Muk151QII28 and ordered from Invitrogen Technologies, USA.

Muk-30 and Muk-33, with sequences

5'-CAGCTGCCTCAGGAGG**GATATCAAACGCATACTCTTG**-3' and

5'-GATTCTCAGCTGCCTCAGGAGGAGG**GATATCAAACGCATACTCTTGAATCATG**-3'

respectively, have an *EcoRV* recognition site, shown in bold. Muk-30 allowed for the deletion of the nucleotide sequence encoding E²²¹ amino acid residue and the substitution of the nucleotide sequence encoding Q²²² amino acid residue with a G residue in the mukwa seed lectin specificity loop. Muk-33 allowed for the substitution of the nucleotide sequence encoding both the E²²¹ and Q²²² residues in the mukwa seed lectin specificity loop with two G residues.

Muk-34, with sequence

5'-GACACCTTCTAT**GCGCA**AAGACTATAACACTTGGGATCCGAATTAC-3', has an *FspI* recognition site, shown in bold. This mutagenic primer allowed for the substitution of the nucleotide sequence encoding S¹³⁷ amino acid residue of the mukwa seed lectin with that of tyrosine.

3.2.7.2 Oligonucleotide-Directed Mutagenesis by Elimination of a Unique Restriction Site

The Unique Site Elimination (U.S.E.) Mutagenesis kit (Amersham Biosciences, 2000) was used for the site-directed *in vitro* mutagenesis experiments. The method utilizes two oligonucleotide primers that introduce mutations into the plasmid. The selection primer carries a mutation that destroys a unique, non-essential restriction endonuclease recognition site in the plasmid. Elimination of this site renders the mutated plasmid DNA resistant to restriction digestion, providing the basis for selection. The desired target mutation is introduced into a defined sequence of the plasmid by the target mutagenic primer.

Clone Muk151QII28, pUC18 carrying the cDNA encoding the wild type mukwa seed lectin, was used as the original template. Two selection primers, designated SELECPUC1 and SELECPUC2, were designed based on the pUC18 nucleotide sequence. SELECPUC1

converts the *NdeI* recognition site to the *NcoI* recognition site while SELECPUC2 converts the *NcoI* recognition site to the *NdeI* recognition site. SELECPUC1 and SELECPUC2, with sequences 5'-GAGTGCACCAT**TGGGCGGTGTGAAAT**-3' and 5'-GAGTGCACCAT**TATGCGGTGTGAAAT**-3' have the respective *NcoI* and *NdeI* restriction endonuclease recognition sites, shown in bold.

3.2.7.3 Introduction of the *EcoRV* Recognition Site into Clone Muk151QII28

Mutagenic primer Muk-17 was used to introduce the *EcoRV* recognition site into clone Muk151QII28, with SELECPUC1 as the selection primer.

3.2.7.3.1 Annealing of Primers to Plasmid DNA and Synthesis of Mutant DNA Strand

Site-specific mutations are introduced into the plasmid under the direction of the selection and the mutagenic primers. The double-stranded target plasmid is first heat-denatured and the primers simultaneously annealed to the same strand of circular single-stranded plasmid DNA. Synthesis of the new strand of DNA results in incorporation of the altered sequences.

For the annealing reaction, 100 ng of plasmid DNA (clone Muk151QII28), 100 ng of selection primer (SELECPUC1), 100 ng of mutagenic primer (Muk-17), 2 µl of 10 x One-Phor-All Buffer PLUS (100 mM Tris acetate, 100 mM magnesium acetate and 500 mM potassium acetate, pH 7.5) and water up to 20 µl were mixed in an Eppendorf tube. The tube was incubated in boiling water for 5 min and immediately chilled on ice for 5 min. The tube was centrifuged briefly and then incubated at room temperature for 30 min.

The mutagenesis reaction mix was set up by adding 7 µl of the nucleotide mix and 3 µl of the reaction mix (containing T4 DNA polymerase and T4 DNA ligase) to the tube. The contents of the tube were mixed gently, the tube centrifuged briefly and incubated at 37°C for 1 hr.

After incubation, the reaction was stopped by heating at 85°C for 15 min after which the tube was centrifuged briefly and placed on ice.

3.2.7.3.2 Primary Selection by Restriction Endonuclease Digestion

The product of the previous procedure in Section 3.2.7.3.1 is a heteroduplex plasmid consisting of a wild type parental strand and a new full-length strand that carries the desired mutation but no longer contains the unique restriction recognition site. This product, however, also contains wild type molecules that were never used as templates of the DNA synthesis primed by the two primers. If this mixed plasmid population is incubated with the restriction enzyme whose recognition site was eliminated by the U.S.E. selection primer, the mutated plasmids are resistant to digestion but the wild type molecules are linearized.

To the tube containing the mutagenesis reaction mix (30 µl), 5 µl of 10 x One-Phor-All Buffer PLUS, 1 µl (10 units) of *NdeI* and 14 µl of water were added and mixed gently. After centrifuging briefly, the reaction tube was incubated at 37°C for 2 hr.

3.2.7.3.3 Transformation of *E. coli* WK6mutS Cells

Transformation efficiency is 10 to 1000-fold greater with circular DNA than with linear DNA (Conley and Saunders, 1984). Consequently, if the mixture of circular heteroduplex DNA and linear wild type DNA is used to transform an *E. coli* strain that is deficient in repair of mismatched bases, transformation with mutant plasmid is favoured.

To enhance mutagenesis efficiency, the bacterial cells were transformed using electroporation. In this technique, the application of a pulse of electricity to bacterial cells is thought to open pores in the membranes through which DNA molecules can pass from the medium into the cytoplasm. This method gives transformation efficiencies greater than available with most chemical methods (Bonnardel *et al.*, 2000).

3.2.7.3.3.1 Preparation of Electrocompetent Cells

A 15% glycerol stock of *E. coli* WK6mutS cells (stored at -80°C) was thawed on ice and a loopful of the cells aseptically streaked onto LB agar. The plate was incubated at 37°C overnight. A single colony was aseptically picked, inoculated into 5 ml of liquid LB broth and grown overnight with shaking at 250 rpm (New Brunswick Innova Shaker, USA). Two millilitres of the overnight culture were inoculated into 200 ml of LB liquid broth in a 500 ml flask and grown at 37°C with shaking at 250 rpm to an absorbance of 0.40-0.45 at 660 nm.

The cell suspension was then chilled on ice for 30 min, centrifuged at 4 000 x g for 15 min at 4°C (Sorvall RC-5B Refrigerated Superspeed Centrifuge, Du Pont Instruments, USA) and the supernatant discarded. The cells were re-suspended in the original culture volume (200 ml) of ice-cold water, centrifuged again at 4 000 x g for 15 min at 4°C and the supernatant discarded. Twice, the cells were re-suspended in 100 ml of ice-cold water followed by centrifugation. The cells were then re-suspended in 4 ml of 10% sterile glycerol and collected by centrifugation. Finally, the cells were re-suspended in 600 µl of 10% sterile glycerol, dispensed into 40 µl aliquots in pre-chilled Eppendorf tubes and stored at -80°C. To assess electrocompetence, the cells were transformed with pUC18 and plated on LB agar supplemented with 100 µg/ml carbenicillin. Untransformed cells were also plated on LB agar containing 100 µg/ml carbenicillin as a negative control.

3.2.7.3.3.2 Electroporation

Electroporation was carried out using the BIO-RAD Gene Pulser (BIO-RAD, USA). The plasmid DNA sample from the mutagenesis reaction was dialyzed on nitrocellulose millipore filter (0.025 µm White VSWP, 13 mm) for 1 hr against sterile water in a Petri dish to remove all salts in preparation for electroporation. Two transformation reactions were set up using 2 µl and 20 µl of the mutagenesis reaction plasmid DNA respectively.

Electrocompetent cells were thawed on ice, gently mixed with the plasmid DNA and incubated on ice for 1 min. The Gene Pulser apparatus was set at 25 μ F, 200 ohms and 2.5 kV. The mixture of cells and plasmid DNA was transferred to an ice-cold 0.1 cm electroporation cuvette and the suspension shaken to the bottom of the cuvette. The cuvette was then placed in a safety chamber slide and pushed until the cuvette was seated between the contacts in the base chamber of the electroporation device.

After pulsing the apparatus once, the cuvette was removed from the chamber and the cells immediately re-suspended in 900 μ l of SOC medium (2% {w/v} Bacto-tryptone, 0.5% {w/v} Bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄ and 20 mM glucose). The cell suspension was transferred to a 17 x 100 mm polypropylene tube and incubated at 37°C for 1 hr with shaking at 225 rpm.

One hundred microlitres of each transformation reaction were plated out on LB agar supplemented with 100 μ g/ml carbenicillin and incubated at 37°C overnight. To each of the remaining transformation reaction, 3 ml of SOC medium supplemented with 100 μ g/ml carbenicillin were added and the cultures grown at 37°C overnight with shaking at 250 rpm.

3.2.7.3.4 Enrichment for Mutant Plasmids

When the circular heteroduplex plasmids replicate, the first round of replication generates a wild type plasmid that carries the original restriction recognition site and a mutated plasmid that does not. If DNA from the first set of transformants is isolated, digested with the same restriction enzyme to linearize the wild type molecules and then used to transform a standard laboratory *E. coli* strain, the proportion of mutant plasmid within the population is increased.

Plasmid DNA from the *E. coli* WK6*mutS* cells was isolated from the overnight cultures using the QIAprep Spin Miniprep kit (Qiagen, 2000a). This DNA was subjected to a second round of restriction enzyme selection. Five microlitres of plasmid DNA were mixed with 2 μ l of 10 x One-Phor-All Buffer PLUS, 1 μ l (10 units) of *Nde*I and 12 μ l of water and incubated at

37°C for 2 hr. After incubation, the plasmid DNA was used to transform *E. coli* K514 cells using Hanahan and co-workers' method (see Section 3.2.2.8). Individual transformant colonies were used to prepare plasmid minipreps and mutants were identified by restriction endonuclease digestion with *EcoRV* and *EcoRI* followed by agarose gel analysis.

3.2.7.4 Introduction of the *BspEI* Recognition Site into Clone Muk151QII28 Carrying the *EcoRV* Recognition Site

Mutagenic primer Muk-18 was used to introduce the *BspEI* recognition site into clone Muk151QII28, now carrying an *EcoRV* recognition site. SELECPUC2 was used as the selection primer. To achieve this, the U.S.E. Mutagenesis kit (Amersham Biosciences, 2000) was used as has already been described in Section 3.2.7.3. Mutants were identified by restriction endonuclease digestion with *BspEI* and *EcoRI* followed by agarose gel analysis. Mutants were also digested with *EcoRV* to check if both restriction endonuclease recognition sites had been successfully incorporated. Sequencing primers Muk-6, that binds upstream, and Muk-12 (sequence 5'-CTTCTTCTGAGCAGTGTAC-3'), that binds downstream of the mutated region, were used to check if the nucleotide sequence was as intended (see Appendix C).

3.2.7.5 Introduction of the UEA II and the ECorL Specificity Loops into the Mukwa Seed Lectin

The nucleotide sequence encoding the mukwa seed lectin specificity loop was removed from clone Muk151QII28 mutants carrying the *EcoRV* and *BspEI* recognition sites by digesting the mutant plasmid DNA with these restriction endonucleases. Equimolar concentrations of primer pairs specifying for the UEA II (Muk-19 and Muk-20) and the ECorL (Muk-21 and Muk-22) specificity loops respectively, were separately annealed and ligated with the *EcoRV*-*BspEI*-digested mutants. It must be noted however, that the introduction of the DNA segment specifying for the ECorL specificity loop into the *EcoRV* recognition site gave

in the last amino acid residue of this segment, ESP10 as aspartic acid (D) instead of the required alanine (A).

To approximately 1 µg of mutant plasmid DNA, 2 µl of 10 x NEB buffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂, 1 mM dithiothreitol, pH 7.9 at 25°C), 1 µl (10 units) *EcoRV*, 1 µl (10 units) *BspEI* and water up to a total volume of 20 µl were added and the mixture incubated at 37°C for 2 hr. A 5 µl aliquot of the digested DNA was run on agarose gel to check for complete digestion.

For the annealing reaction, 5 µl each of a 5 µM solution of Muk-19 and Muk-20 were mixed with 10 µl of water, incubated in boiling water for 5 min and then allowed to slowly cool down (by switching off the water bath) to about 40°C. Similarly Muk-21 and Muk-22 were also annealed. For the ligation reaction, 1 µl (1.25 pmol) of annealed primers, 5 µl (0.1 pmol) of digested plasmid DNA, 2 µl of 10 x T4 DNA ligase buffer, 0.1 µl (0.5 units) of T4 DNA ligase and water to a total volume of 20 µl were mixed and incubated at 16°C overnight. The whole ligation mixture was used to transform *E. coli* K514 cells using Hanahan and co-workers' method (see Section 3.2.2.8).

Colony screening PCR using primer pairs Muk-6/Muk-20 and Muk-6/Muk-22 was used to identify pools carrying putative mutants with the UEA II and the ECorL specificity loops respectively (see Section 3.2.2.9).

Cultures of individual colonies were then grown and mutants positively identified by digesting with *BstZI* for the UEA II loop D and with *AgeI* for the ECorL loop D. Sequencing primers Muk-6 and Muk-12 were used to confirm the sequence of the mutants.

3.2.7.6 Replacement of Aspartic Acid Residue with Alanine (ESP10) in Mutants Carrying the ECorL Specificity Loop

The nucleotide sequence coding for the D residue, the tenth amino acid residue of the ECorL specificity loop had to be changed to give that of the required A residue (ESP10). To achieve this, the U.S.E. Mutagenesis kit (Amersham Biosciences, 2000) was used, with

mutagenic primer Muk-35 and selection primer SELECPUC1, as has already been described in Section 3.2.7.3. Colony screening PCR using primers Muk-6 and pUC RP (sequence 5'-AGCGGATAACATTTTCACACAGGA-3') was carried out as described in section 3.2.2.9 and pools carrying putative mutants identified by digesting the amplified product with *Apa*LI. Plasmid DNA was isolated from cultures of individual colonies and mutants positively identified by digestion with *Apa*LI followed by sequencing using primers Muk-6 and Muk-12.

3.2.7.7 Substitution of Glutamic Acid (E²²¹) and Glutamine (Q²²²) Residues in the Mukwa Seed Lectin Specificity Loop by One or Two Glycine (G) Residues

Mutagenic primer Muk-30 was used to delete the nucleotide sequence encoding E²²¹ amino acid residue and substitute the Q²²² amino acid residue with a G residue in the mukwa seed lectin specificity loop. Similarly, mutagenic primer Muk-33 was used to substitute the nucleotide sequence encoding both the E²²¹ and the Q²²² amino acid residues with G residues. For both mutants, mutagenic primer Muk-34 was used to substitute the nucleotide sequence encoding S¹³⁷ of the mukwa seed lectin with that of tyrosine. To achieve this, the U.S.E. Mutagenesis kit (Amersham Biosciences, 2000) with clone Muk151QII28 as the template and SELECPUC1 as the selection primer, was used for both reactions.

For each mutagenesis reaction, 100 ng of each of the three primers (the selection primer and the two mutagenic primers), were added to the same reaction tube and annealed at the same time. Mutants were identified by digesting with *Eco*RV and *Eco*RI as well as with *Fsp*I followed by sequencing. Sequencing was carried out using primers Muk-6 that binds upstream and Muk-12 that binds downstream of the mutated region.

3.2.7.8 Sub-Cloning of the DNA Encoding the Mukwa Mutant Lectins into the pBADMycHisA-Wild Type Mukwa Seed Lectin Clone

Clone Muk151QII28 mutants were used as template DNA to amplify, by PCR, the DNA fragments carrying the incorporated mutations. The primers used were Muk-6 that binds

upstream and Muk-25 that binds downstream of the mutated region. The PCR-generated fragments were then sub-cloned into the pBADMycHisA-wild type mukwa seed lectin clone.

The presence of unique recognition sites, for restriction endonucleases *ClaI* and *AvrII*, in the DNA coding for the mukwa seed lectin which are absent in both pUC18 and pBADMycHisA vectors, enabled the sub-cloning (see Appendix C). By digesting with these enzymes, the wild type mukwa seed DNA segment in the pBADMycHisA recombinant clone could be excised and be replaced by the corresponding DNA segment from the pUC18 recombinant mutants that had been generated by PCR.

The PCR-generated fragments (amplified under the same conditions as has already been described in Section 3.2.1.2) and the pBADMycHisA-wild type mukwa seed lectin clone were digested with different combinations of *ClaI*, *AvrII* and *EcoRI*. The digested DNA was ligated using T4 DNA ligase. The expectation was that upon ligation, some of the wild type mukwa seed DNA would be replaced by the mutated DNA segment. Colony screening PCR was then set up using primers pBAD1 and Muk-25. Digesting with the restriction endonucleases specific for each mutant, followed by sequencing using primers pBAD1, Muk-6 and Muk-25, led to the identification of mutants.

3.2.7.9 Expression of the Mukwa Seed Mutant Lectins

The mukwa seed mutant lectins were expressed in *E. coli* TOP10 cells as has been described in Section 3.2.4.2 for the wild type mukwa seed lectin.

The pBADMycHisA-mukwa seed mutant lectins were transformed into *E. coli* TOP10 cells as described in Section 3.2.2.8 and selected on LB agar containing 100 µg/ml carbenicillin. Varying amounts of L-arabinose over a ten-thousand-fold range, from 0.00002% to 0.2% (w/v), were used to determine the appropriate amount of L-arabinose required for maximum expression of the mukwa seed mutant lectins. To analyze the expression, SDS-PAGE was then used as described in Section 3.2.4.3.

3.2.8 Characterization of the Mukwa Seed Mutant Lectins

After introducing specific mutations into the DNA coding for a given protein, changes in the functional properties of the protein can be studied. Following expression, the desired protein must be isolated and characterized.

3.2.8.1 Assaying for Cross-Reactivity of the Mutant Lectins with Antiserum Directed Against the Mukwa Seed Lectin

In order to determine if the mutant lectins could be purified by affinity chromatography on an antibody column, the mutant lectins had to be assayed for cross-reactivity with antiserum directed against the mukwa seed lectin. The crude extract from *E. coli* cells expressing the mutant lectins was first subjected to ammonium sulphate precipitation and the different fractions used in agarose gel immunodiffusion tests.

3.2.8.1.1 Ammonium Sulphate Fractionation

In the presence of high salt concentrations, proteins tend to aggregate and precipitate out of solution. Many contaminants such as nucleic acids and polysaccharides remain in solution and hence salting out of proteins helps in the purification process. To fractionate the protein sample, small quantities of the salt, to give the desired percent saturation, are added at a time with stirring until all the salt has dissolved and then the precipitated protein is removed by centrifugation.

Fifty millilitre cultures of each of the bacterial cells expressing the wild type lectin and the mutant lectins were grown and the respective crude cell extracts subjected to ammonium sulphate fractionation. The 2 ml crude cell extract was made up to 5 ml with saline azide solution and subjected to 30% saturated ammonium sulphate precipitation (176 g of salt/ 1 litre of solution) according to Green and Hughes (1955). The precipitated protein was removed by centrifugation at 15 000 rpm (SS34 Rotor) for 15 min and dissolved in 1 ml of saline azide.

The supernatant was subjected to 60% ammonium sulphate saturation (198 g of salt/ litre of solution) and the precipitated protein dissolved in 1 ml of saline azide. The supernatant was further subjected to 90% ammonium sulphate saturation (227 g of salt/ litre of solution) and the precipitated protein dissolved in 1 ml of saline azide. The 1 ml protein fractions were dialyzed overnight against 2 x 2 L of saline azide. After dialysis, the dialysates were centrifuged in a microfuge at 13 000 rpm for 15 min to remove any protein precipitates, and the clear supernatant was used in the immunodiffusion tests.

3.2.8.1.2 Agarose Gel Immunodiffusion Tests

The basic principle of agarose gel immunodiffusions is the formation of complexes that precipitate out of solution. The procedure involves adding antigen and antibody to wells in an agarose gel and allowing them to diffuse into the agarose gel. Precipitin lines form where the antigen meets specific antibodies.

Six microscope slides were mounted onto a specially designed tray (Gelman Instruments, USA). In order for the agarose to stick onto the slides, a 0.2% (w/v) agar solution (in water) supplemented with glycerol (1 drop/25 ml) was applied onto the slides using a paintbrush. When the solution had solidified, 10 ml of 1% (w/v) agarose solution (in 50 mM barbital buffer, pH 8.6 containing 0.02% {w/v} NaN_3) supplemented with 3% (w/v) PEG-6 000, were applied to each half of the tray and allowed to solidify. Holes were then punched into the agarose using a gel punch (Gelman Instruments, USA) to give the shape of a rosette, and the agarose removed using a Pasteur pipette attached to a vacuum pump.

Five microlitres of the antiserum directed against the mukwa seed lectin were applied into the central hole of each rosette on the slides and 5 μl of each of the ammonium sulphate fractions of the recombinant lectins applied into the outer holes. Pre-immune serum was included as a negative control and purified native mukwa seed lectin as the positive control. The tray was then placed on top of several layers of wetted tissue paper in a microscope slide tank, incubated at 4°C overnight, after which the slides were observed for precipitin lines.

3.2.8.2 Assaying for Recognition of Glycoproteins by the Mukwa Seed Mutant Lectins

In order to determine if the mukwa seed mutant lectins were capable of binding any glycoproteins, different glycoproteins were blotted onto nitrocellulose, incubated with the ammonium sulphate fractions containing the recombinant lectins (determined from the agarose gel immunodiffusion tests) and immunoprobed with antiserum directed against the mukwa seed lectin.

3.2.8.2.1 Preparation of Glycoproteins of High Variability

Porcine and bovine plasma powders are useful sources of glycoproteins of high variability (Nollet *et al.*, 1999). However, plasma powders contain huge amounts of serum albumin, a non-glycosylated protein (Jordan and Goldstein, 1994; Guzman-Partida *et al.*, 2004), and this must be removed first.

Five grams of each of porcine and bovine plasma powders (Functional Animal Proteins, Vepro Products, Belgium) were respectively suspended in 30 ml of 20 mM phosphate buffered saline (PBS), pH 7.2 containing 0.02% (w/v) NaN₃ and dialyzed overnight against 2 x 1 L of the same buffer. The dialyzed plasma suspensions were centrifuged at 15 000 rpm (SS34 rotor) for 15 min and the pellet discarded. The supernatant was subjected to 40% ammonium sulphate saturation (243 g of salt/ litre of solution) and the precipitated protein (the IgG fraction) dissolved in 10 ml of 20 mM PBS, pH 7.2. The supernatant was dialyzed overnight against 2 x 1 L of 25 mM Na-acetate buffer, pH 5.2.

The dialysate was centrifuged at 15 000 rpm for 15 min to remove any precipitated protein and the supernatant applied onto a DEAE-Sepharose column (2 x 18 cm) that had been equilibrated in 25 mM Na-acetate buffer, pH 5.2. After the unbound fraction had been collected, the BSA fraction was eluted with 25 mM Na-acetate buffer, pH 4.4. The rest of the protein was eluted with 150 mM Na-acetate buffer, pH 4.0. The different fractions, including the IgG fractions were dialyzed overnight against 2 x 2 L of 20 mM PBS, pH 7.2.

Half of each fraction was dialyzed for 4 hr against 2 x 2 L of double distilled water and freeze-dried (Virtis, USA) overnight.

3.2.8.2.2 Analysis of Binding of Glycoproteins by Recombinant Lectins

Five milligrams of each of the freeze-dried fractions, as well as lyophilized hen egg white, were dissolved in 1 ml 1 x SDS-PAGE sample solution (Laemmli, 1970) and the quality and quantity checked by SDS-PAGE as has already been described in Section 3.2.4.3. Five microlitres of each sample were then used for Western blotting as has also been described in Section 3.2.4.4. An additional step was, however, included after blocking. This step involved incubating the blot for 5 hr in a 1 in 100 dilution of the 30-60% ammonium sulphate fraction (250 μ l in 25 ml) of the recombinant lectin. Each of the different recombinant lectins was analysed separately. The recombinant lectin solution was prepared in 3% BSA (w/v), 0.01% Tween-20 (v/v) in 10 mM PBS, pH 7.3. After washing the blots, antiserum directed against the mukwa seed lectin was added and the protocol continued as before.

Negative controls, in which sugar moieties were first destroyed by incubating the blots in 10 mM NaIO₄ in 0.2 M Na-acetate buffer, pH 5.0 overnight before adding the recombinant lectins, were included. As the sodium periodate reaction must take place in the dark, the tray containing the blots was covered by aluminium foil.

For the wild type recombinant lectin, a third Western blot processed in the presence of 100 mM methyl- α -D-mannopyranoside was included. The presence of this sugar in the lectin sample and the washing solution would inhibit the lectin from binding to mannose moieties on the glycoproteins.

3.2.9 Purification of the Mukwa Seed Mutant Lectins

The mukwa seed mutant lectins were purified to determine their physico-chemical characteristics and to produce their respective X-ray crystals. The cross-reaction of these mutant lectins with antibodies directed against the native mukwa seed lectin (determined from the agarose gel immunodiffusions) meant that these recombinant lectins could be isolated by affinity chromatography on a mukwa seed lectin antibody-Sepharose-4B column. Consequently, antiserum directed against the native mukwa seed lectin had to be developed and specific antibodies isolated from total IgG by affinity chromatography on a mukwa seed lectin-Sepharose-4B column.

3.2.9.1 Immunogen Preparation

The native mukwa seed lectin, used as the immunogen, was purified from the seed crude extract on Sephadex G-75. The lectin interacts reversibly with Sephadex and is eluted slower than the rest of the protein in the extract (Manyumwa, 1998).

The mukwa seed coat was peeled off using a sharp blade and the seeds ground to a fine powder using a mortar and pestle. The seed meal was defatted by stirring in hexane (5 ml/g) at room temperature for 2 hr. The hexane was removed by filtration under vacuum and the defatted seed meal dried in the fumehood for 1 hr. The resulting dry seed meal was then extracted with saline azide (5 ml/g) on a magnetic stirrer at 4°C overnight. The crude extract was filtered through cheesecloth and clarified by centrifugation at 15 000 rpm (SS34 rotor) at 4°C for 15 min.

The crude extract was subjected to 30% ammonium sulphate saturation followed by centrifugation at 15 000 rpm for 15 min. The resulting supernatant was subjected to 60% ammonium sulphate saturation and the pellet obtained dissolved in saline azide (1 ml/g of dry seed meal).

Five millilitres of the 30-60% ammonium sulphate fraction of the mukwa seed extract was applied on a 2.5 x 28 cm Sephadex G-75 column and eluted with saline azide at a flow rate of 0.34 ml/min. Fractions of the fourth (retarded) peak were analysed by SDS-PAGE to check for purity, pooled together and concentrated using a Millipore Ultrafree-15 Centrifugal Filter device of 10 000 MW limit (Sigma, UK). The lectin concentration was determined by measuring the OD₂₈₀, using 1.33 as the specific absorbance of the lectin.

3.2.9.2 Immunization and Bleeding of Rabbits

When an immunogen is administered to an animal, it is recognized as foreign and the animal produces antibodies directed against it. Repeated immunization (boosting) increases the amount and the affinity of antibody produced. Rabbits are often chosen for immunization purposes as they are small enough to keep in an indoor facility and yet are large enough to produce reasonable amounts (up to 25 ml of blood per bleed) of antiserum.

Before immunization, 1 ml of pre-immune serum was collected from each of the three rabbits. The inoculum was prepared by mixing the mukwa seed lectin with an equal volume of Freund's adjuvant and vortexing vigorously. A total of 500 µg of the immunogen in 2 ml was injected into each rabbit in four sites, two intramuscular and two intradermal/subcutaneous. Complete adjuvant was used for the initial immunization while the incomplete adjuvant was used for the subsequent boosts, once every 2 weeks.

After 8 weeks, the rabbits were bled once every week from the ear by restraining the animal and making a small cut in one of the veins in the outer edge of the ear. The blood was left to clot at room temperature for 1 hr and then stored at 4°C overnight to allow the clot to retract. The antiserum was obtained by centrifugation at 2 000 x g and used in agarose gel diffusions to test for the presence of specific antibodies.

3.2.9.3 Isolation of Total IgG by Caprylic Acid Fractionation and DEAE-Cellulose Chromatography

IgG is the major serum antibody and constitutes about a sixth of the protein present. Caprylic acid precipitates contaminating proteins and leaves the antibody in solution. The product, which is about 50% IgG, is contaminated with albumin and other serum proteins (Steinbuch and Audran, 1969). Contaminants are removed from the IgG by ion-exchange chromatography using the anion exchanger diethylaminoethyl (DEAE)-cellulose. The separation is on the basis of charge and makes use of the fact that IgG has a higher isoelectric point than other serum proteins.

A measured volume (50 ml) of antiserum was placed in a beaker and mixed with two volumes (100 ml) of 0.06 M acetate buffer, pH 4.0. Caprylic acid, 7.5 ml per original 100 ml of antiserum, was added dropwise with stirring, following which the mixture was stirred for a further 30 min to ensure maximum precipitation. The precipitate was removed by centrifugation at 2 000 x g at room temperature and the supernatant filtered through Whatman filter paper to remove fines and through 0.22 µm filters to clear. The clear supernatant, now called the caprylic acid cut, was concentrated to reduce the volume to about 50 ml. Caprylic acid was removed by dialyzing against 2 x 2 L of 0.02 M phosphate buffer pH 6.2 containing 0.02% (w/v) sodium azide at 4°C.

A 2.5 x 20 cm DEAE-cellulose column was equilibrated with 0.02 M phosphate buffer pH 6.2 at a flow rate of 1.0 ml/min using a pump with absorbance of the eluent being monitored at 280 nm. Twenty-five millilitres of the caprylic acid cut was loaded onto the column and the IgG eluted in the void volume until the absorbance at 280 nm approached the set base line. The bound contaminants were eluted with 0.02 M phosphate buffer pH 6.2 containing 2 M NaCl. The column was regenerated by washing with the starting buffer containing 0.02% (w/v) sodium azide and stored at room temperature in the same buffer. Protein recoveries were determined using the BCA protein assay.

3.2.9.4 Immobilization of the Native Mukwa Seed Lectin onto Sepharose-4B

Glycoproteins can be coupled to Sepharose beads that have been activated with cyanogen bromide (CNBr). Cyanogen bromide reacts with the hydroxyl groups of Sepharose to form imidocarbonate and carbamate groups. During the subsequent coupling of the glycoprotein to the active product, the imidocarbonate groups react with amino groups belonging to the glycoprotein with the formation of stable covalent linkages (Kohn and Wilchek, 1984; Beeckmans, 1999).

Forty-five milligrams of the native mukwa seed lectin (see Section 3.2.9.1) in a volume of 20 ml were dialyzed against 3 x 2 L of coupling buffer, 0.1 M NaHCO₃ buffer, pH 8.2 containing 0.5 M NaCl and immobilized onto 4 g of CNBr-activated Sepharose-4B. The Sepharose was washed and re-swelled on a sintered glass filter using 800 ml of 1 mM HCl followed by 1 L of coupling buffer and the lectin solution added immediately. The lectin-gel suspension was mixed in an end-over-end mixer for 4 hr at room temperature and then overnight at 4°C. The gel was collected by filtration and the lectin content in the filtrate compared with the original lectin content by measuring absorbance at 280 nm.

Remaining active groups were blocked by incubating the gel in 40 ml of 0.2 M glycine, pH 8.0 at room temperature for 1 hr with shaking. Adsorbed protein was washed away with 1 L of coupling buffer followed by 1 L of 0.1 acetate buffer, pH 4.0 containing 0.5 M NaCl followed by 1 L coupling buffer. Thereafter the gel was stored in coupling buffer containing 0.02% (w/v) sodium azide.

3.2.9.5 Isolation of Specific IgG on the Mukwa Lectin-Sepharose-4B Column

Specific anti-mukwa seed lectin IgG can be isolated from total IgG by affinity chromatography on a mukwa seed lectin-Sepharose-4B column. However, as mannose moieties are found on IgG molecules, the saccharide must be added to both the sample and running buffer to inhibit the recognition of the specific IgG by their mannose moieties.

Ten millilitres of the total IgG fraction containing 0.2 M α -D-mannose (36 g/L) was applied onto the mukwa seed lectin-Sepharose-4B column (1.5 x 6 cm) equilibrated with PBS, pH 7.3 containing 0.2 M mannose. The unbound protein was washed with a large volume of PBS, pH 7.3 containing 0.2 M mannose until the absorbance at 280 nm approached the set base line. The bound fraction was eluted with 0.1 M citrate acid buffer, pH 3.0. The peak (eluted protein) was immediately dialyzed overnight at 4°C against 2 x 2 L of PBS, pH 7.3. The column was regenerated by washing with PBS, pH 7.3 containing 0.02% NaN₃ and stored at room temperature in the same solution.

3.2.9.6 Immobilization of Specific IgG Directed Against the Mukwa Seed Lectin onto Sepharose-4B

The specific IgG directed against the mukwa seed lectin was concentrated using a Millipore Ultrafree-15 Centrifugal Filter device of 10 000 MW limit (Sigma, UK) to give a volume of about 20 ml (approximately 1 mg/ml) and then dialyzed overnight against 3 x 2 L of 0.1 M NaHCO₃ buffer, pH 8.2 containing 0.5 M NaCl. The IgG was coupled to CNBr-activated Sepharose-4B beads as has already been described in Section 3.2.9.4.

3.2.9.7 Isolation of the Mukwa Seed Recombinant Lectins on the Anti-Mukwa Seed Lectin IgG-Sepharose-4B Column

Ten times 50 ml cultures of *E. coli* cells harbouring plasmids with the cDNA encoding each recombinant lectin were separately grown and induced with 20% L-arabinose as has already been described for the wild type lectin (see Section 3.2.6.4). After sonication, the crude extract was applied onto the column (1.5 x 4.5 cm) equilibrated with saline azide supplemented with 5 mM Ca²⁺ and 5 mM Mn²⁺ ions. The bound lectins were eluted with 0.1 M acetic acid directly into 3 ml aliquots of PBS, pH 7.3 up to a maximum volume of 6 ml. The peak (eluted protein) was dialyzed overnight at 4°C against 2 x 2 L of PBS, pH 7.3. The column was regenerated by washing with PBS, pH 7.3 containing 0.02% (w/v) NaN₃ and

stored at room temperature in the same solution. Protein recoveries were determined using the BCA assay, and the purification process for each recombinant lectin analysed by SDS-PAGE.

3.2.10 Assaying the Mukwa Seed Mutant Lectins for Activity

The biological activity of the mukwa seed mutant lectins was assessed by haemagglutinating assays. Initially, the 30-60% ammonium sulphate fractions of the crude extract from *E. coli* cells expressing the wild type and the mutant lectins were used to assay for agglutination using untreated human erythrocytes.

Rabbit, sheep, bovine, goat and porcine blood were collected and the erythrocytes washed as described in Section 3.2.5.3.1. Each of the pure mutant lectin was concentrated to approximately 1 mg/ml and haemagglutinating tests carried out using the different animal erythrocytes as described in Section 3.2.5.3.2. Both the native mukwa seed lectin and the wild type recombinant lectin were included as controls.

4. RESULTS

4.1 Expression of the cDNA Encoding the Mukwa Seed Lectin in *E. coli*

4.1.1 Amplification of the cDNA Encoding the Mature Wild Type Mukwa Seed Lectin by PCR

Clone Muk151QII28, a pUC18 vector carrying the cDNA encoding the wild type mukwa seed lectin, was used as template DNA to amplify the DNA region that codes for the mature mukwa seed lectin. Amplification by PCR was carried out using each of primers Muk-23 and Muk-24 with Muk-25 respectively. Both primer pairs gave amplification products of the expected size, approximately 700 bp, for all the template dilutions used. No amplification product was obtained for the negative control where template DNA had not been included.

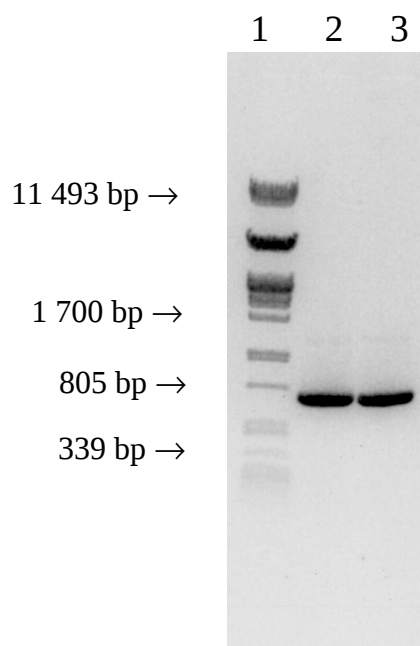


Figure 4.1: PCR analysis of clone Muk151QII28. Resolution is on a 0.8% (w/v) agarose gel and shows the PCR amplification products of the mature mukwa seed lectin cDNA. Lane 1 shows the Lambda *Pst*I MM size marker. Lanes 2 and 3 show the amplification products generated using primer pairs Muk-24/Muk-25 and Muk-23/Muk-25 respectively.

4.1.2 Cloning of the cDNA Encoding the Mature Wild Type Mukwa Seed Lectin into the Expression Vector pBADMycHisA

The expression vector, pBADMycHisA was digested with *Nco*I, filled in with T4 DNA polymerase and then digested with a second restriction endonuclease, *Eco*RI. The amplified DNA region coding for the mature wild type seed lectin was also digested with *Eco*RI and cloned into the digested pBADMycHisA.

The number of bacterial colonies obtained for putative recombinants carrying the insert generated using PCR primers Muk-23 and Muk-25 was approximately a tenth of that obtained for putative recombinants carrying the insert generated using PCR primers Muk-24 and Muk-25. The lower number of colonies showed that the bacterial cells harbouring the former construct may have failed to grow due to suspected lethal effects to these cells.

For each construct, all the pools of bacterial colonies screened for possible recombinants gave amplification products of the expected size, about 700 bp and the negative controls gave no amplification product. One pool was selected for each construct and plasmid DNA isolated from individual colonies. Upon digestion of the plasmid DNA with *Eco*RI and *Eco*RV, putative recombinants gave the expected two fragments of approximately 3.3 kb and 1.5 kb (see Figure 4.2). This restriction pattern confirmed that a DNA fragment approximately 700 bp had been cloned as the empty expression vector would only be 4.1 kb in size. Putative recombinants with inserts generated using the PCR primer pair Muk-24/Muk-25 were designated pGV4976 and pGV4977 while those with inserts generated using the primer pair Muk-23/Muk-25 were designated pGV4978 and pGV4979 respectively.

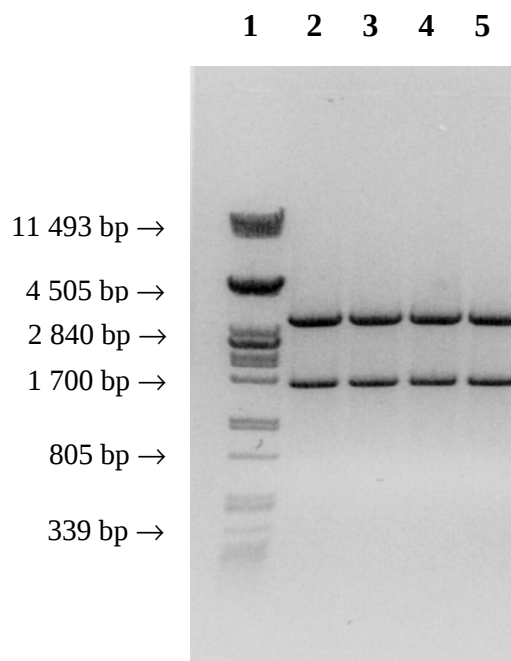


Figure 4.2: *EcoRI-EcoRV* restriction pattern of pBADMycHisA putative clones carrying the mature mukwa seed lectin 700 bp DNA fragment generated by PCR. Resolution is on a 0.8% (w/v) agarose gel. Lane 1 shows the Lambda *PstI* MM size marker. Lanes 2 and 3 show the respective DNA fragments of the putative clones pGV4976 and pGV4977 generated by PCR using primers Muk-24 and Muk-25. Lanes 4 and 5 show the respective DNA fragments of the putative clones pGV4978 and pGV4979 generated by PCR using primers Muk-23 and Muk-25.

4.1.3 Sequencing of the Mature Wild Type Mukwa Seed Lectin cDNA Clones

The pBADMycHisA-wild type mukwa seed lectin clones were sequenced using the pBAD Forward Primer (pBAD1) (Invitrogen Corporation, 2000) particularly to check if the reading frame of the expression vector had been maintained at the point of ligation. The nucleotide sequence showed that only those clones carrying the PCR fragment generated using primers Muk-24 and Muk-25, that is, pGV4976 and pGV4977 were in frame (see Figure 4.3). These clones carried the insert in which the codon for the first amino acid, glutamine, of the mature lectin had been excluded.

5'-GGAGGAATTA**ACCATGG**ATTCCCTTTCCTTCGGCTTCCC-3'

Figure 4.3: Nucleotide sequence around the *Nco*I-filled-in recognition site of pBAD*MycHisA* clones pGV4976 and pGV4977 carrying the insert generated using PCR primers Muk-24 and Muk-25. The region showing the filled-in *Nco*I recognition site is shown in bold.

Clones carrying the DNA fragment generated using PCR primers Muk-23 and Muk-25, pGV4978 and pGV4979, had the last nucleotide of the expression vector missing and were therefore out of frame (see Figure 4.4). Consequently, these clones were not used for expression.

5'-GGAGGAATTA**ACCATCA**AGATTCCCTTTCCTTCGGCTTCCC-3'

Figure 4.4: Nucleotide sequence around the *Nco*I-filled-in recognition site of pBAD*MycHisA* clones pGV4978 and pGV4979 carrying the insert generated using PCR primers Muk-23 and Muk-25. The filled-in *Nco*I recognition site, shown in bold, has a G missing.

The cloned fragments were then sequenced entirely using pBAD1, Muk-6 and Muk-25 and this confirmed that no errors had occurred during amplification by PCR (see Figure 4.5). Subsequently, pGV4977 was used for the expression experiments.

1	M D S L S F G F P T F P S D Q K	
Muk151QII28	- ATG GATTCCCTTTCTTCGGCTTCCCTACGTTTCCTTCAGACCAGAAAA	73
17	N L I F Q G D A Q I K N N A V Q L	
Muk151QII28	ACCTAATCTTCCAAGGTGATGCACAGATTAACCAACGACGTCCAGCTT	123
34	T K T D S N G N P V A S T V G R I	
Muk151QII28	ACCAAGACAGACAGTAATGGTAACCCAGTGGCAAGCACTGTTGGG CGA AT	173
51	L F S A Q V H L W E K S S S R V	
Muk151QII28	CTTATTCTCGGCGCAAGTGCACCTTTGGGAAAAAGCTCAAGC AG AGTGG	223
67	A N F Q S Q F S F S L K S P L S N	
Muk151QII28	CAAACCTTTCAATCCCAGTTCAGCTTTTCCCTCAAATCACCGCTTTCCAAT	273
84	G A D G I A F F I A P P D T T I P	
Muk151QII28	GGAGCCGACGGCATTGCCTTCTTCATCGCACCGCCGGATACTACCATACC	323
101	S G S G G G L L G L F A P G T A	
Muk151QII28	GAGCGGTTCTGGGCGGAGGTCTCCTAGGGCTCTTTGCACCTGGAAGTGTCTC	373
117	Q N T S A N Q V I A V E F D T F Y	
Muk151QII28	AAAACACGTCCGCGAACCAAGTGATTGCGGTTGAGTTTGACACCTTCTAT	423
134	A Q D S N T W D P N Y P H I G I D	
Muk151QII28	GCTCAAGACTCTAACACTTGGGATCCGAATTACCCGCACATTGGAATCGA	473
151	V N S I R S V K T V K W D R R D	
Muk151QII28	TGTCAACTCCATT AGG TCTGTGAAGACTGTGAAATGGGAT AGGAG AGATG	523
167	G Q S L N V L V T F N P S T R N L	
Muk151QII28	GTCAATCCCTTAATGTCCTCGTAACATTTAACCTTAGCACTCGAAACCTG	573
184	D V V A T Y S D G T R Y E V S Y E	
Muk151QII28	GATGTCGTTGCTACTTACTCTGATGGTACA AG ATACGAGGTGTCTTATGA	623
201	V D V R S V L P E W V R V G F S	
Muk151QII28	GGTTGACGTG AGG TCCGTGCTTCCAGAATGGGTT AG AGTTGGATTCTCAG	673
217	A A S G E Q Y Q T H T L E S W S F	
Muk151QII28	CTGCCTCAGGAGAGCAATATCAAACGCATACTCTTGAATCATGGTCTTTC	723
234	T S T L L Y T A	
Muk151QII28	ACCTCAACCTTGTTGTACACTGCT TAG -----	750

Figure 4.5: Nucleotide and amino acid sequence of the mature mukwa seed lectin. The first and last amino acids, glutamine residues, have been replaced by methionine and a stop codon respectively. Codons shown in bold specify for the amino acid arginine.

4.1.4 Analysis of Expression of the cDNA Encoding the Wild Type Mukwa Seed Lectin in *E. coli*

For expression, the pBADMycHisA-wild type mukwa seed lectin clone pGV4977 was transformed into *E. coli* TOP10 cells. Sequencing had revealed the presence of seven arginine codons AGA/AGG that are rarely used in *E. coli* (see Figure 4.5). Consequently, in a separate expression experiment, the host *E. coli* TOP10 cells were transformed with both plasmids pGV4977 and BL21-CodonPlus-RIL, which enhances expression of these codons. Varying amounts of L-arabinose over a ten thousand-fold range, from 0.00002% to 0.2% (w/v), were then used to determine the appropriate amount of L-arabinose required for maximum expression of the wild type mukwa seed lectin.

Analysis of expression was carried out using SDS-PAGE and Western blotting. The SDS-PAGE gave the expected band of approximately 25 kDa in size which increased in intensity with increasing concentration of L-arabinose as can be seen in Figure 4.6. Figure 4.7 shows that addition of plasmid BL21-CodonPlus-RIL did not significantly enhance expression of the mukwa seed lectin. On the contrary, a very pronounced *E. coli* protein band, approximately 22 kDa can be seen. Subsequently, plasmid BL21-CodonPlus-RIL was not included in the expression experiments.

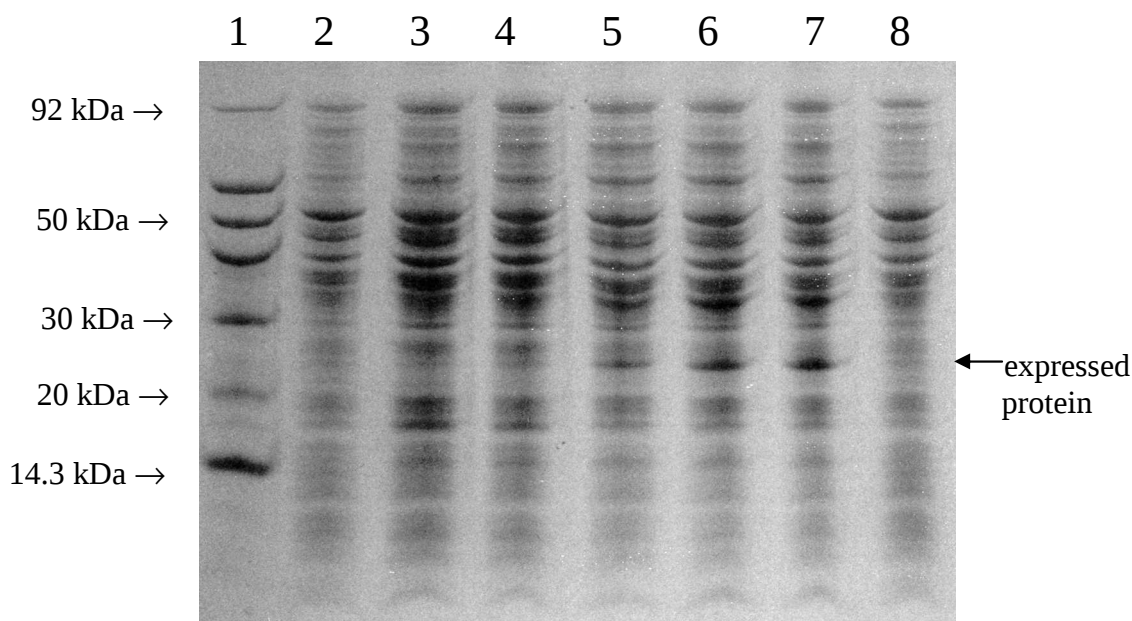


Figure 4.6: SDS-PAGE analysis of total protein in *E. coli* TOP10 cells transformed with pGV4977, a pBADMyHisA clone carrying the cDNA encoding the mature wild type mukwa seed lectin. Lane 1 shows the protein MM size markers. Lanes 2 and 8 show the total protein before induction of expression. Lanes 3 to 7 show the total protein obtained after induction with ten-fold increment of L-arabinose, from 0.00002% to 0.2% (w/v).

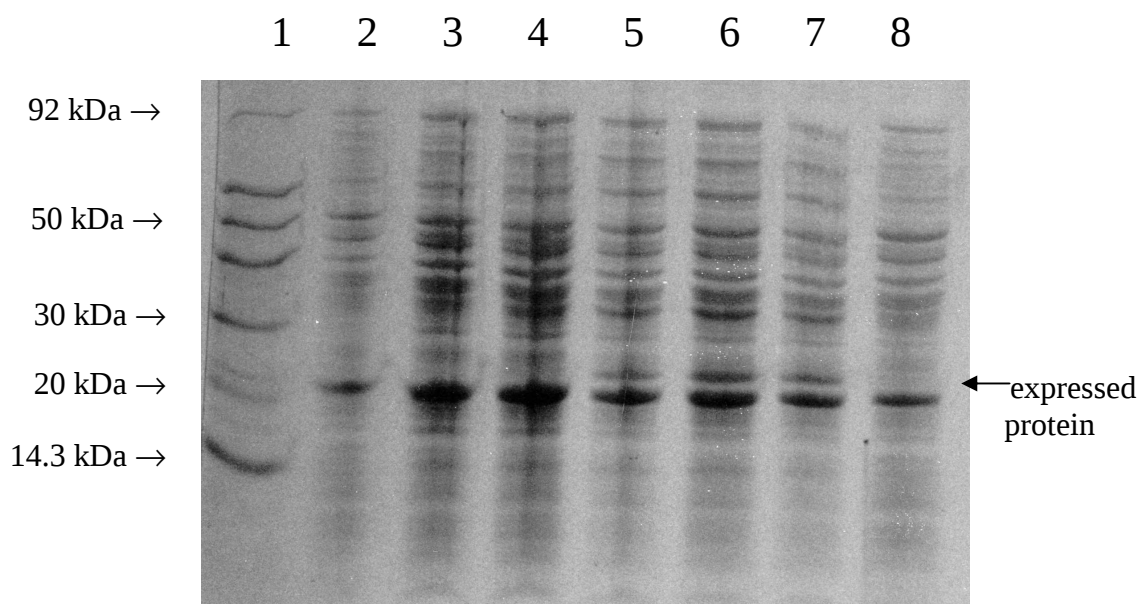


Figure 4.7: SDS-PAGE analysis of total protein in *E. coli* TOP10 cells transformed with both BL21CodonPlus-RIL and pGV4977, a pBADMyHisA clone carrying the cDNA encoding the mature wild type mukwa seed lectin. Lane 1 shows the protein MM size markers. Lanes 2 and 8 show the total protein before induction of expression. Lanes 3 to 7 show the total protein obtained after induction with ten-fold increment of L-arabinose, from 0.00002% to 0.2% (w/v).

Western blotting showed that the expressed protein was recognized by antiserum directed against the mukwa seed lectin (see Figure 4.8). As for SDS-PAGE, the expressed protein band increased in intensity with increasing concentration of L-arabinose. However, with increasing concentration of the expressed lectin, three distinct protein bands of lower molecular mass than the lectin were also recognized by the antiserum. These bands probably correspond to degradation products and possible aggregation of these products could be responsible for the smear observed in lanes 6 to 8 in Figure 4.8. Degradation also seemed to increase with the age of the sample.

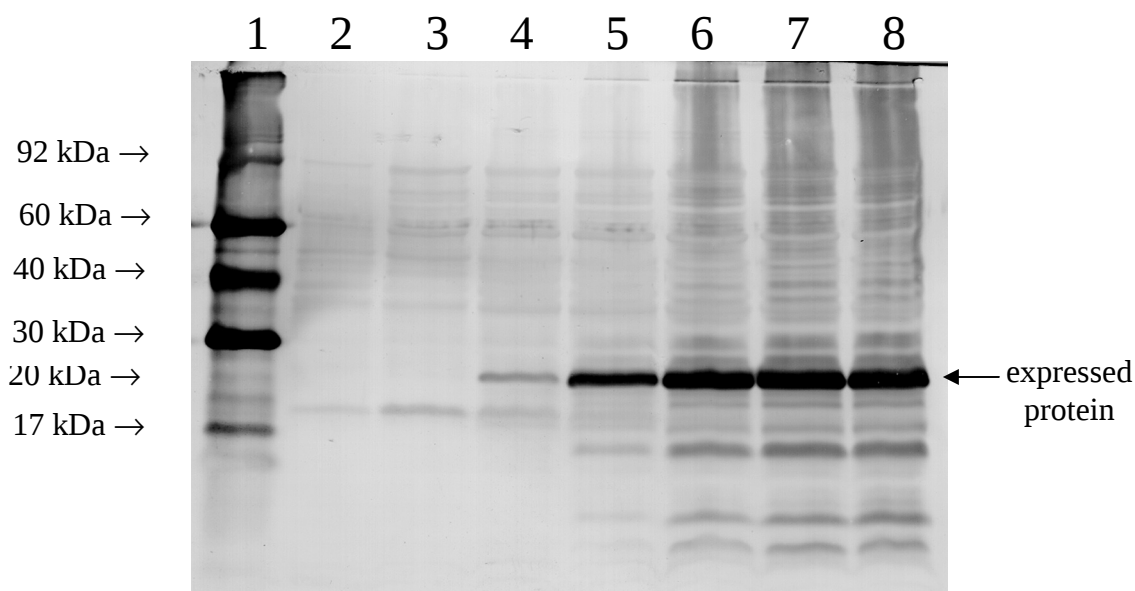


Figure 4.8: Western blot analysis of total protein in *E. coli* TOP10 cells transformed with pGV4977, a pBADMyHisA clone carrying the cDNA encoding the mature wild type mukwa seed lectin. Immunoprobings were done using rabbit antiserum directed against the mukwa seed lectin followed by protein A-biotin and extravidin phosphatase. Lane 1 shows the biotinylated protein MM size markers. Lanes 2 shows the total protein from cells transformed with an empty pBADMyHisA while lane 3 shows the total protein before induction of expression. Lanes 4 to 8 show the total protein obtained after induction with ten-fold increment of L-arabinose, from 0.00002% to 0.2% (w/v).

4.1.5 Assaying the Expressed Wild Type Mukwa Seed Lectin for Activity

Cells carrying the expressed protein were re-suspended in saline azide supplemented with 5 mM CaCl_2 and 5 mM MnCl_2 , 1/25 volume of the original culture. The cell suspension was sonicated, centrifuged and the supernatant used to test for haemagglutination and inhibition of agglutination by methyl- α -D-mannopyranoside. In all these experiments, the native lectin, isolated from mukwa seeds was used as a control.

Starting from a 50 ml culture, the expressed protein gave 32 agglutinating units with rabbit erythrocytes and did not agglutinate any of the untreated human blood groups A1+, A2+, B+, AB+ and O+. Agglutination of rabbit erythrocytes by the expressed lectin was inhibited by methyl- α -D-mannopyranoside just like the native lectin.

4.1.6 Purification of the Expressed Wild Type Mukwa Seed Lectin by Affinity Chromatography

From a 1 x 500 ml culture showing 8 agglutinating units and a 10 x 50 ml culture showing 16 agglutinating units, about 3.8 mg and 7.7 mg of the lectin could be purified by affinity chromatography on mannose-Sepharose-4B respectively. The lectin was eluted with 0.3 M mannose in saline azide. The difference in yield is most likely due to better aeration obtained in small volumes compared to large volumes. Preparation of the cell extract in the absence or in the presence of a protease inhibitor cocktail (1 mM PMSF + 2 μ M leupeptin + 2 μ M pepstatin A) made no difference in the yield of the expressed lectin. The purified lectin had a slightly lower molecular mass than the native lectin and degradation products observed in the Western blot in Figure 4.8 can also be seen in Figure 4.9. Like the native lectin isolated from mukwa seeds, the purified lectin had a specific absorbance of $\text{OD}_{280 \text{ nm}}$ 1 mg/ml = 1.3 and showed an absorbance ratio of $\text{OD}_{280 \text{ nm}} / \text{OD}_{250 \text{ nm}} \sim 3$.

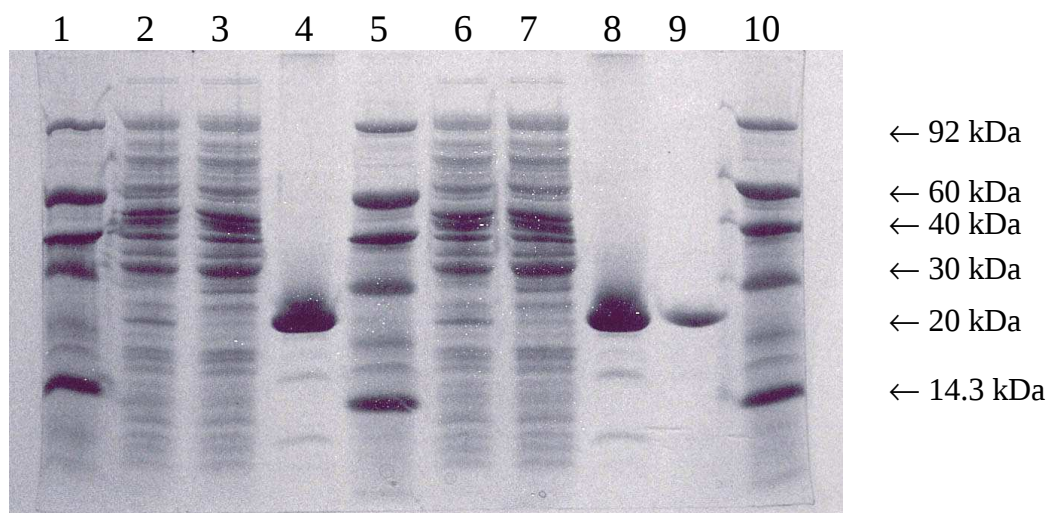


Figure 4.9: SDS-PAGE analysis of the purified wild type mukwa seed lectin expressed in *E. coli* TOP10 cells. Purification, by affinity chromatography on mannose-Sepharose-4B, was performed either in the presence (lanes 2-4) or in the absence (lanes 6-8) of a protease inhibitor cocktail (1 mM PMSF + 2 μ M leupeptin + 2 μ M pepstatin A). Lanes 1, 5 and 10 show the protein MM size markers. Lanes 2 and 6 show the extract from *E. coli* cells expressing the wild type mukwa seed lectin. Lanes 3 and 7 show the unbound fraction. Lanes 4 and 8 show the lectin fraction, eluted with 0.3 M mannose in saline azide. Lane 9, the positive control, shows the native lectin isolated from mukwa seeds.

4.2 Mutagenesis of the Sugar-Binding Specificity Loop of the Mukwa Seed Lectin

4.2.1 Introduction of Restriction Endonuclease Recognition Sites into Clone Muk151QII28 for the Removal of the Mukwa Seed Lectin Specificity Loop

Mutagenic primers Muk-17 and Muk-18 were used with the unique site elimination mutagenesis kit (Amersham Biosciences, 2000) to introduce the *EcoRV* and *BspEI* restriction endonuclease recognition sites respectively into clone Muk151QII28.

After the introduction of the *EcoRV* recognition site into clone Muk151QII28, 12 colonies were randomly picked and screened for possible mutants. When the plasmid DNA from these colonies was digested with *EcoRI* and *EcoRV*, four putative mutants designated Muk151QII28-*EcoRV*-Mut1 through Muk151QII28-*EcoRV*-Mut4, were identified. Of these, Muk151QII28-*EcoRV*-Mut1 and Muk151QII28-*EcoRV*-Mut4 were used as template DNA for

the introduction of the *BspEI* recognition site. Sixteen bacterial colonies were randomly picked and screened for possible mutants. Digestion of the corresponding plasmid DNA with *EcoRI* and *BspEI* led to the identification of three putative mutants, designated pGV4980, pGV4981 and pGV4982 respectively. These putative mutants gave the expected two bands of approximately 3 200 bp and 300 bp upon digestion with *BspEI* and *EcoRI*, and only one band of approximately 3 500 bp upon digestion with *EcoRV* (see Figure 4.10). This restriction pattern and sequencing confirmed that both the *BspEI* and the *EcoRV* recognition sites had successfully been introduced into clone Muk151QII28 (see Figure 4.11).

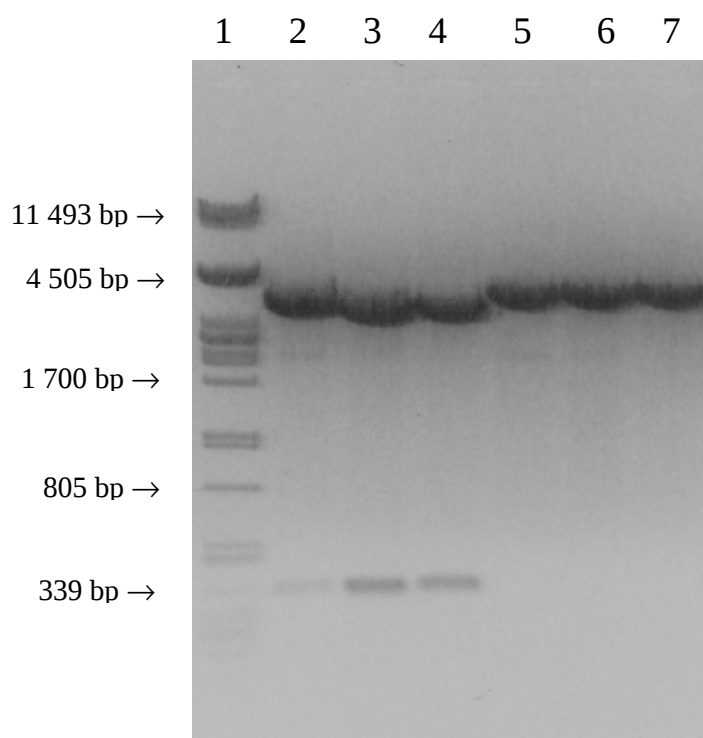


Figure 4.10: Restriction pattern of putative mutants of clone Muk151QII28 carrying the *BspEI* and *EcoRV* recognition sites. Resolution is on a 0.8% (w/v) agarose gel. Lane 1 shows the Lambda *PstI* MM size marker. Lanes 2, 3 and 4 show the respective DNA fragments of putative mutants pGV4980, pGV4981 and pGV4982 digested with *BspEI* and *EcoRI*. Lanes 5, 6 and 7 show the DNA fragment of the same putative mutants digested with *EcoRV*.



Figure 4.11: Nucleotide and amino acid sequence of the mukwa seed lectin mutant carrying the *BspEI* and *EcoRV* recognition sites. The region of the specificity loop targeted for mutagenesis is underlined and mutations are shown in bold.

4.2.2 Introduction of the UEA II and ECorL Specificity Loops into the Mukwa Seed Lectin

The mukwa seed lectin specificity loop was removed by digesting pGV4981, a Muk151QII28 mutant carrying the *EcoRV* and *BspEI* recognition sites with these restriction endonucleases. Equimolar concentrations of the complementary primers Muk-19 and Muk-20 were annealed and these specified for the specificity loop of the *Ulex europaeus* II lectin (UEA II).

Similarly, equimolar concentrations of the complementary primers Muk-21 and Muk-22 were also annealed and these specified for the specificity loop of the *Erythrina corallodendron* lectin (ECorL). These annealed oligonucleotides were then ligated with the

BspEI-EcoRV-digested Muk151QII28 mutant, pGV4981. However, introduction of the ECorL loop into the *EcoRV* recognition site would give the amino acid residue ESP10 as D instead of the required A (see Figure 3.2).

After transformation, for each reaction, all the six pools of bacterial colonies screened for possible recombinants gave amplification products of approximately 400 bp, the expected size. Both negative controls gave no amplification products. For each reaction, cultures of individual colonies corresponding to the pool that gave the most intense band were grown and plasmid DNA isolated. For the first reaction, digestion of plasmid DNA with *BstZI* and *EcoRI* led to the identification of three putative recombinants carrying the UEA II specificity loop. These putative mutants, designated pGV4988, pGV4989 and pGV4990, although incompletely digested, gave the two expected band sizes of approximately 3 200 bp and 300 bp. Similarly, for the second reaction, putative recombinants carrying the ECorL specificity loop, with the amino acid residue ESP10 as D instead of A, also gave the same restriction pattern upon digestion with *AgeI* and *EcoRI* (see Figure 4.12). These four putative mutants were designated pGV4991, pGV4992, pGV4993 and pGV4994 respectively.

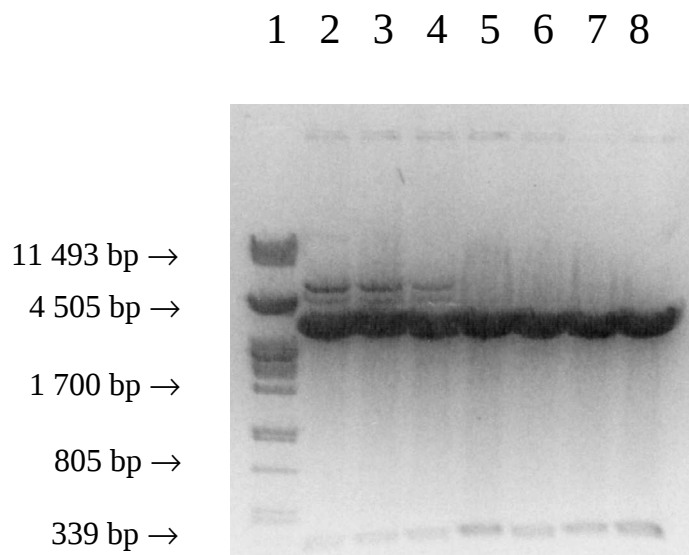


Figure 4.12: Restriction pattern of putative mutants of clone Muk151QII28 with the respective UEA II specificity loop and the ECorL specificity loop. The ECorL specificity loop has amino acid residue ESP10 as D instead of A. Resolution is on a 0.8% (w/v) agarose gel. Lane 1 shows the Lambda *PstI* MM size marker. Lanes 2, 3 and 4 show the respective DNA fragments of putative mutants pGV4988, pGV4989 and pGV4990 digested with *BstZI* and *EcoRI*. Lanes 5, 6, 7 and 8 show the respective DNA fragments of putative mutants pGV4991, pGV4992, pGV4993 and pGV4994 digested with *AgeI* and *EcoRI*.

Sequencing confirmed that the nucleotide sequence was as desired. It must, however, be noted that introduction of the UEA II specificity loop into the *EcoRV* recognition site resulted in the phenylalanine residue USP9 being substituted for by tyrosine, another aromatic residue in the UEA II mutant. Introduction of the ECorL specificity loop into the *EcoRV* recognition site gave the amino acid residue ESP10 as D instead of the required A (see Figure 4.13).

		I A P P D T T I P	
Muk151QII28	-----	CATCGCACCGCCGGATACTAC CATACC	323
		Muk-6 →	
101		S G S G G G L L G L F A P G T A	
Muk151QII28		GAGCGGTTTCGGGCGGAGGTCTCTAGGGCTCTTTGCACCTGGAAGTCTC	373
117		Q N T S A N Q V I A V E F D T F Y	
Muk151QII28		AAAACACGTCCGCGAACCAAGTGATTGCGGTTGAGTTTGACACCTTCTAT	423
134		A Q D S N T W D P N Y P H I G I D	
Muk151QII28		GCTCAAGACTCTAACACTTGGGATCCGAATTACCCGCACATTGGAATCGA	473
151		V N S I R S V K T V K W D R R D	
Muk151QII28		TGTCAACTCCATTAGGTCTGTGAAGACTGTGAAATGGGATAGGAGAGATG	523
167		G Q S L N V L V T F N P S T R N L	
Muk151QII28		GTCAATCCCTTAATGTCCTCGTAACATTTAACCTTAGCACTCGAAACCTG	573
184		D V V A T Y S D G T R Y E V S Y E	
Muk151QII28		GATGTCGTTGCTACTTACTCTGATGGTACAAGATACGAGGTGTCTTATGA	623
201		V D V R S V L P E W V R V G F S	
Muk151QII28		GGTTGACGTGAGGTCCGTGCTTCCAGAATGGGTTAGAGTTGGATTCTCCG	673
217		G A T G A Q R D A D Q T H T L E S	
Muk151QII28		GAGCGACCGGTGCGCAGCGTGATGCGGATCAAACGCATACTCTTGAATCA	723
		AgeI	
234		W S F T S T L L Y T A Q K K	
Muk151QII28		TGGTCTTTCACCTCAACCTT GTTGTACACTGCTCAGAAGAAG -----	765
		← Muk-12	

Figure 4.13: Nucleotide and amino acid sequence of the mukwa seed lectin mutant carrying the ECorL specificity loop with amino acid residue ESP10 as D instead of the required A. The ECorL specificity loop amino acid residues are shown in bold and ESP10 is bordered.

4.2.3 Replacement of Amino Acid Residue D with A (ESP10) in Mutants Carrying the ECorL Specificity Loop

Ligation of the annealed primers Muk-21 and Muk-22 to the *BspEI-EcoRV*-Muk151QII28 mutants gave mutants with the nucleotide sequence coding for the amino

residue ESP10 as D instead of A. This D residue, in mutants pGV4991 and pGV4993 was replaced with the appropriate A residue using mutagenic primer Muk-35 and the unique site elimination mutagenesis kit.

Using primers Muk-6 and pUC RP, an amplification product of about 650 bp, the expected size, was obtained for all the 20 pools of bacterial colonies screened as well as for the controls in which pGV4991 and pGV4993 had been used as template DNA. Digestion of the amplification products with *ApaLI* showed that putative mutants were present in 14 pools. Cultures of individual colonies from two selected pools were grown and plasmid DNA isolated. Digestion of the plasmid DNA with *ApaLI* led to the identification of six putative mutants (see Figure 4.14). These putative mutants gave a unique band of approximately 900 bp that was absent in the control pGV4991. Two of the putative mutants were designated pGV5014 and pGV5015 and their respective restriction pattern can be seen in lanes 3 and 8 of Figure 4.14.

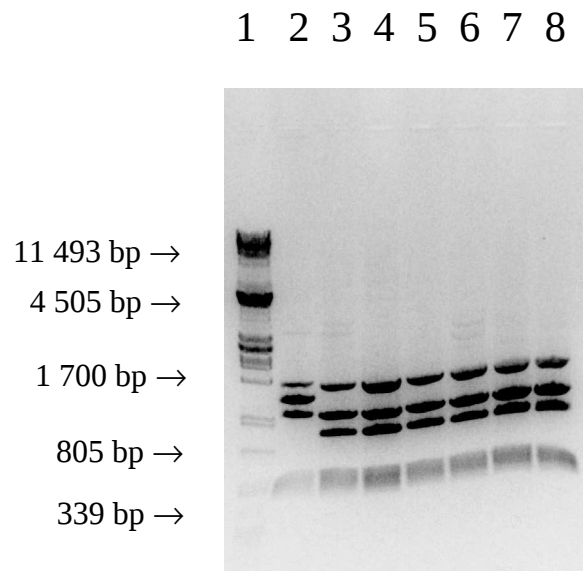


Figure 4.14: *ApaLI*-restriction pattern of putative mutants of clone Muk151QII28 carrying the ECorL specificity loop. Resolution is on a 0.8% (w/v) agarose gel. Lane 1 shows the Lambda *PstI* MM size marker. Lane 2 shows the DNA fragments of pGV4991, a mutant with the ECorL specificity loop with the amino acid residue ESP10 as D instead of A. Lanes 3, 4, 5, 6, 7 and 8 show the respective DNA fragments of the putative mutants.

4.2.4 Substitution of Glutamic Acid (E²²¹) and Glutamine (Q²²²) Residues in the Mukwa Seed Lectin Specificity Loop by One or by Two Glycine Residues

The unique site elimination mutagenesis kit was used to substitute the E²²¹ and Q²²² residues in the specificity loop of the mukwa seed lectin with one or two G residues. Mutagenic primer Muk-30 was used to delete one amino acid and substitute the other by a G residue whereas mutagenic primer Muk-33 was used to substitute both amino acids with G residues. In both mutagenesis reactions, S¹³⁷ was replaced with a tyrosine residue using mutagenic primer Muk-34 and the unique site elimination mutagenesis kit.

After the mutagenesis, twenty colonies from each reaction were randomly screened for possible mutants. Clone Muk151QII28 was included as a control. Upon digestion of the plasmid DNA with *EcoRI* and *EcoRV*, putative mutants in which the E²²¹ and Q²²² residues had been substituted for by one G or two G residues gave the two expected band sizes of approximately 3200 bp and 300 bp (see Figure 4.15). Two mutants, designated pGV5005 and pGV5006, were identified for the S¹³⁷ → Y + E²²¹Q²²² → single G substitution and their respective restriction pattern in lanes 2 and 3 of Figure 4.15 shows incomplete digestion. For the S¹³⁷ → Y + E²²¹Q²²² → double G substitution, three mutants, designated pGV5008, pGV5009 and pGV5010 were identified. The deletion of one amino acid and substitution of the other by a G residue, and the substitution of both amino acids by two G residues respectively, were confirmed by sequencing. Digestion of these putative mutants with *FspI* showed that S¹³⁷, had successfully been substituted for by tyrosine (see Figure 4.16), and this was confirmed by sequencing.

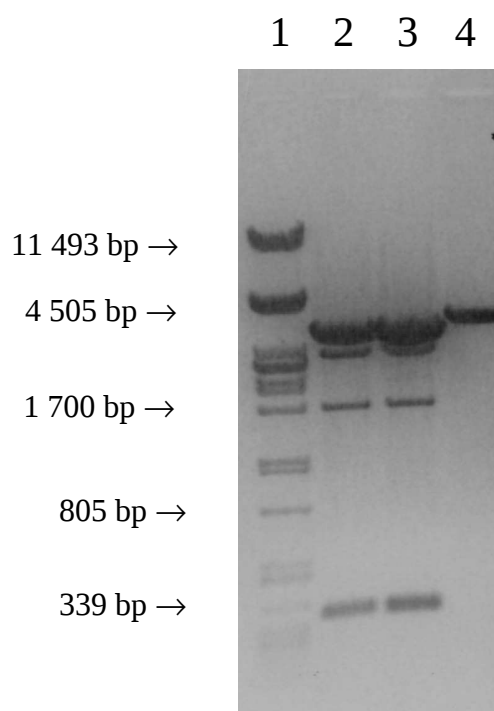


Figure 4.15: *EcoRI-EcoRV*-restriction pattern of putative mutants of clone Muk151QII28 carrying the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution in the specificity loop. Resolution is on a 0.8% (w/v) agarose gel. Lane 1 shows the Lambda *PstI* MM size marker. Lanes 2 and 3 show the respective DNA fragments of putative mutants pGV5005 and pGV5006. Lane 4 shows clone Muk151QII28 DNA fragments.

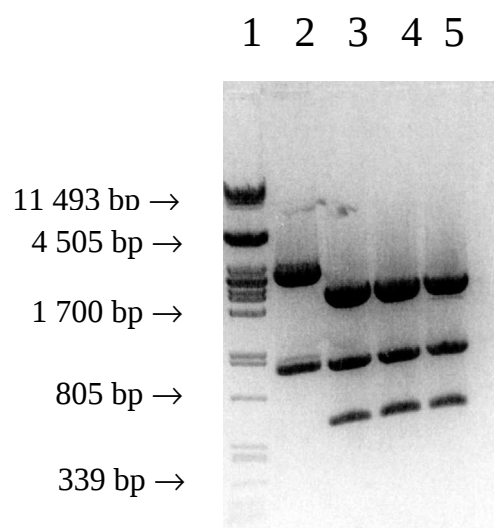


Figure 4.16: *FspI*-restriction pattern of putative mutants of clone Muk151QII28 carrying the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution in the specificity loop. Resolution is on a 0.8% (w/v) agarose gel. Lane 1 shows the Lambda *PstI* MM size marker. Lane 2 shows clone Muk151QII28 DNA fragments. Lanes 3, 4 and 5 show the respective DNA fragments of the putative mutants pGV5008, pGV5009 and pGV5010.

4.3 Cloning of the Mukwa Seed Lectin Mutants into the Expression Vector pBADMycHisA

Restriction endonucleases *ClaI* and *AvrII* have unique recognition sites in the DNA coding for the mukwa seed lectin which are absent in both pUC18 and pBADMycHisA vectors. The DNA region carrying the mutations was amplified by PCR using primers Muk-6 and Muk-25. Four clones pGV4990, pGV5014, pGV5005 and pGV5008 were used as the respective template DNA for the UEA II specificity loop, ECorL specificity loop, $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution and $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution mutants. All the four representative mutants gave amplification products of about 450 bp, the expected size.

The amplification products and clone pGV4977 (the expression vector pBADMycHisA carrying the DNA coding for the wild type lectin), were digested with different combinations of *ClaI*, *AvrII* and *EcoRI*. For pGV4990, a UEA II specificity loop mutant and pGV5014, an ECorL specificity loop mutant, the amplification products and pGV4977 were digested with *ClaI* and *EcoRI*. For pGV5005, an $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution mutant and pGV5008, an $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution mutant, the amplification products and were digested with the three enzyme pairs *ClaI-AvrII*, *AvrII-EcoRI* and *ClaI-EcoRI*. Clone pGV4977 was correspondingly digested with the respective enzyme pairs. The *ClaI-AvrII* digestion allowed for the $S^{137} \rightarrow Y$ change only. The *AvrII-EcoRI* digestion allowed for both the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G or double G substitutions. The *ClaI-EcoRI* digestion allowed for the $E^{221}Q^{222} \rightarrow$ single G or double G substitutions only (see Appendix C).

Upon ligation, some of the pGV4977 wild type DNA was replaced with the mutant DNA. For each transformation reaction, all the six pools screened for possible recombinants using primers pBAD1 and Muk-25 gave amplification products of about 750 bp, the expected size. Amplification by PCR using pGV4977 as template DNA also gave a fragment of the same size. Upon digestion of the amplification products with the restriction endonuclease specific for each respective mutant, the restriction pattern showed that mutants were present in all the pools screened.

Individual colonies from selected pools were cultured and plasmid DNA isolated. Digestion of the corresponding plasmid DNA with the restriction endonuclease specific for respective mutants led to the identification of putative mutants. The pBADMyHisA clone carrying the wild type mukwa seed lectin, pGV4977 was used as a control in all cases.

Putative mutants carrying the UEA II specificity loop were linearized when digested with *Bst*ZI as is shown in Figure 4.17. This one band was as expected as the pBADMyHisA-wild type mukwa seed lectin clone pGV4977 does not have a recognition site for *Bst*ZI. Two of the putative mutants were designated pGV5031 and pGV5032 and their restriction pattern is shown in lanes 3 and 5 of Figure 4.17.

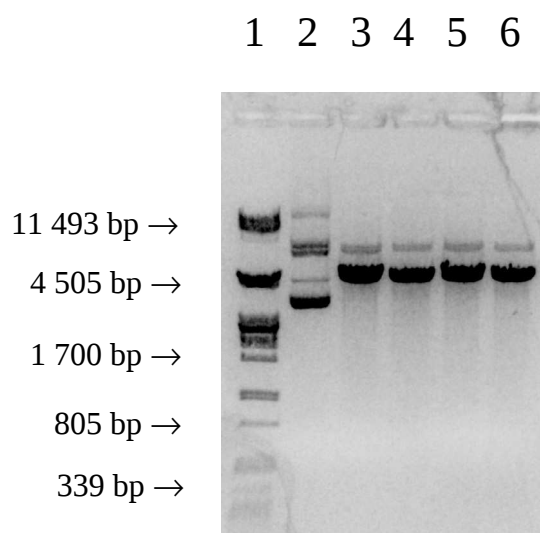


Figure 4.17: *Bst*Z1-restriction pattern of pBADMyHisA-mukwa seed lectin putative mutants carrying the UEA II specificity loop. Resolution is on a 0.8% (w/v) agarose gel. Lane 1 shows the Lambda *Pst*I MM size marker. Lane 2 shows the DNA fragments of pGV4977, the pBADMyHisA clone carrying the cDNA encoding the mature wild type mukwa seed lectin. Lanes 3, 4, 5 and 6 show the respective DNA fragment of the putative mutants.

Digestion of putative mutants carrying the ECorL specificity loop with *Apa*LI gave a unique band of approximately 800 bp that was absent in the control (see Figure 4.18). The respective restriction patterns of three of the putative mutants designated pGV5022, pGV5023 and pGV5024 is shown in lanes 3, 5 and 6 of Figure 4.18. The restriction pattern of the putative mutant shown in lane 4 of Figure 4.18 was not as expected and hence it was not selected for sequencing.

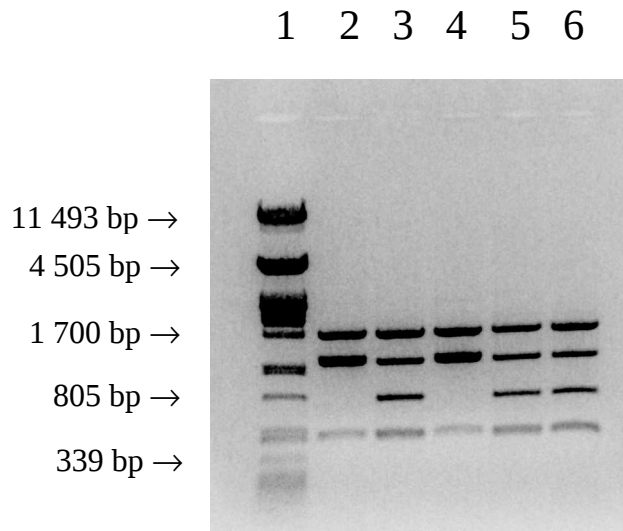


Figure 4.18: *Apa*L1-restriction pattern of pBADMycHisA-mukwa seed lectin putative mutants carrying the ECorL specificity loop. Resolution is on a 0.8% (w/v) agarose gel. Lane 1 shows the Lambda *Pst*I MM size marker. Lane 2 shows the DNA fragments of pGV4977, the pBADMycHisA clone carrying the cDNA encoding the mature wild type mukwa seed lectin. Lanes 3, 4, 5 and 6 show the respective DNA fragments of the putative mutants.

Upon digestion with *Eco*RV and *Fsp*I, putative mutants carrying both the S^{137} to Y change and the $E^{221}Q^{222}$ to G or to GG substitution in the specificity loop gave two DNA fragments of the expected sizes (see Figures 4.19 and 4.20). Putative mutants carrying the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution were designated pGV5018 and pGV5019 while those carrying the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution were designated pGV5025, pGV5026, pGV5027 and pGV5028 respectively. Those mutants that carried only the S^{137} to Y change gave two fragments when digested with *Fsp*I but only one fragment when digested with *Eco*RV. These putative mutants were designated pGV5016 and pGV5017 respectively. On the other hand, mutants carrying only the $E^{221}Q^{222}$ to G or to GG substitution in the specificity loop gave one fragment when digested with *Fsp*I and two fragments when digested with *Eco*RV. Putative mutants carrying the $E^{221}Q^{222} \rightarrow$ single G substitution only were designated pGV5020 and pGV5021 while those carrying the $E^{221}Q^{222} \rightarrow$ double G substitution only were designated pGV5029 and pGV5030 respectively. The restriction pattern of these additional mutants carrying a single change is shown in the same figures (Figures 4.19 and 4.20).

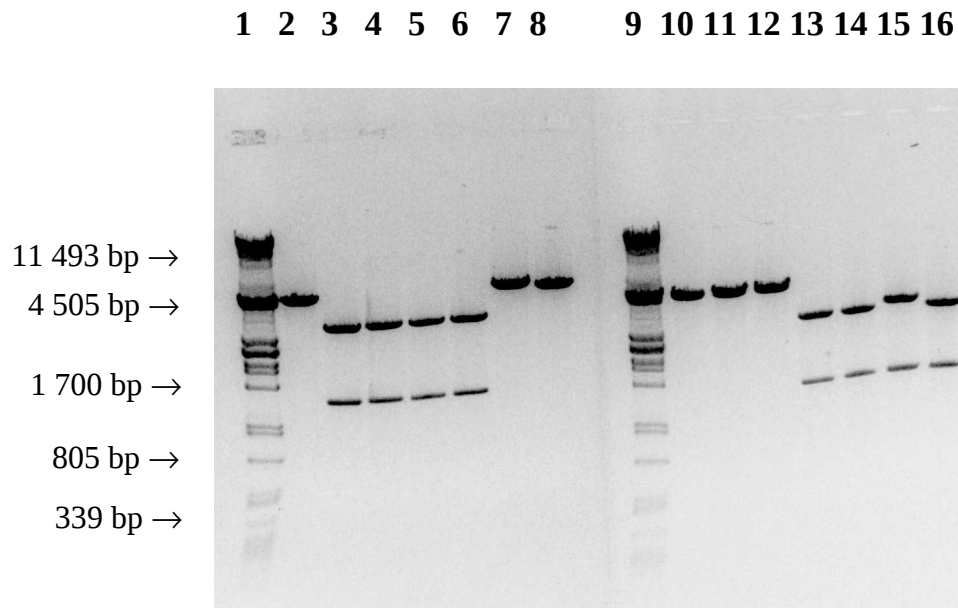


Figure 4.19: Restriction pattern of pBADMyHisA-mukwa seed lectin putative mutants carrying the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution in the specificity loop. Resolution is on a 0.8% (w/v) agarose gel. Lanes 1 and 9 show the Lambda *Pst*I MM size marker. Lanes 2 and 10 show the DNA fragments of pGV4977, the pBADMyHisA clone carrying the cDNA encoding the mature wild type mukwa seed lectin. Lanes 3, 4, 5, 6, 7 and 8 show the respective DNA fragments of the putative mutants pGV5016, pGV5017, pGV5018, pGV5019, pGV5020 and pGV5021 digested with *Fsp*I. Lanes 11, 12, 13, 14, 15 and 16 show the respective DNA fragments of the same putative mutants digested with *Eco*RV.

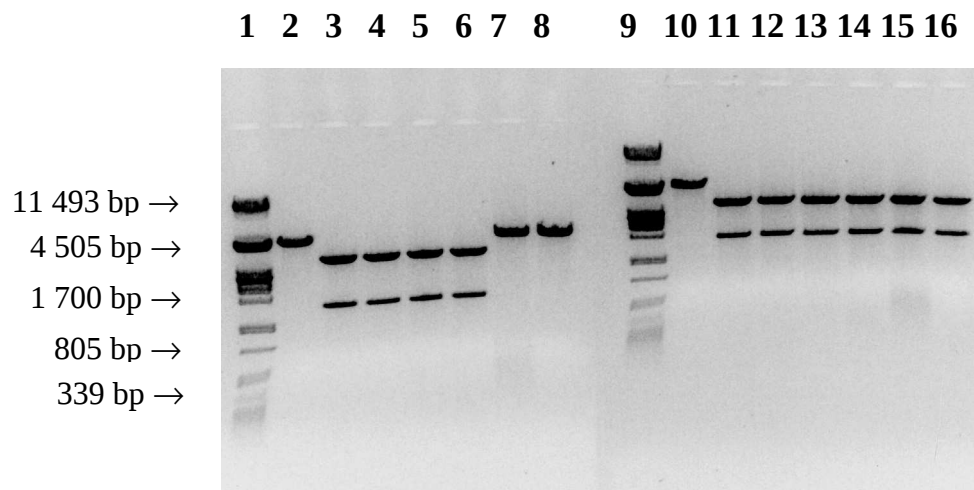


Figure 4.20: Restriction pattern of pBADMyHisA-mukwa seed lectin putative mutants carrying the $S^{137} \rightarrow Y$ and $E^{221}Q^{222} \rightarrow$ double G substitution in the specificity loop. Resolution is on a 0.8% (w/v) agarose gel. Lanes 1 and 9 show the Lambda *Pst*I MM size marker. Lanes 2 and 10 show the DNA fragments of the pBADMyHisA clone carrying the cDNA encoding the mature wild type mukwa seed lectin. Lanes 3, 4, 5, 6, 7 and 8 show the respective DNA fragments of the putative mutants pGV5025, pGV5026, pGV5027, pGV5028, pGV5029 and pGV5030 digested with *Fsp*I. Lanes 11, 12, 13, 14, 15 and 16 show the respective DNA fragments of the same putative mutants digested with *Eco*RV.

4.4 Sequencing of the Mukwa Seed Lectin Mutants

The pBADMyHisA-mukwa seed lectin mutants were sequenced in the region that had been amplified by PCR using primers Muk-6 and Muk-25. Sequencing confirmed that only desired mutations had been incorporated into the DNA fragment coding for the mukwa seed lectin except in pGV5020. The restriction pattern of this putative mutant shown in lanes 7 and 15 of Figure 4.19 also shows that it is slightly bigger than expected. Subsequently pGV5031, pGV5022, pGV5018 and pGV5026 representing the UEA II specificity loop, ECorL specificity loop, $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution and $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution mutants respectively were selected for further analysis (see Figures 4.21 to 4.24). Additional mutants pGV5016, pGV5021 and pGV5030 representing the $S^{137} \rightarrow Y$ change only, $E^{221}Q^{222} \rightarrow$ single G change only and $E^{221}Q^{222} \rightarrow$ double G change only were selected as controls.

The sequences in both Figures 4.21 and 4.22 show that the codon for S^{216} was changed from the TCA found in the wild type mukwa seed lectin to TCC due to the introduction of the *BspEI* recognition site. From the sequence of the mukwa seed lectin mutant carrying the UEA II specificity loop shown in Figure 4.21, it can be seen that the amino acid residue USP9 is tyrosine instead of the phenylalanine that is found in the specificity loop of the native UEA II. The sequence of the mukwa seed lectin mutant carrying the ECorL specificity loop shown in Figure 4.22, gives amino acid residues ESP1 to ESP10 exactly as they are in native ECorL.

In Figures 4.23 and 4.24, the codon for A^{134} is given as GCG instead of the GCT found in the wild type mukwa seed lectin due to the introduction of the *FspI* recognition site. In both sequences, the codon for S^{137} TCT was successfully changed to the intended TAT, the codon for tyrosine. The sequence in Figure 4.23 shows that the amino acid residue E^{221} was successfully deleted and the amino acid residue Q^{222} substituted for by a G amino acid residue. The sequence in Figure 4.24 shows that G amino acid residues successfully replaced both amino acid residues E^{221} and Q^{222} .

1	M D S L S F G F P T F P S D Q K	
Muk151QII28	- ATG GATTCCCTTTCTTCGGCTTCCCTACGTTTCCTTCAGACCAGAAAA	73
17	N L I F Q G D A Q I K N N A V Q L	
Muk151QII28	ACCTAATCTTCCAAGGTGATGCACAGATTAAAAACAACGCAGTCCAGCTT	123
34	T K T D S N G N P V A S T V G R I	
Muk151QII28	ACCAAGACAGACAGTAATGGTAACCCAGTGGCAAGCACTGTTGGGCGAAT	173
51	L F S A Q V H L W E K S S S R V	
Muk151QII28	CTTATTCTCGGCGCAAGTGCACCTTTGGGAAAAAAGCTCAAGCAGAGTGG	223
67	A N F Q S Q F S F S L K S P L S N	
Muk151QII28	CAAACCTTTCAATCCCAGTTCAGCTTTTCCCTCAAATCACCGCTTTCCAAT	273
84	G A D G I A F F I A P P D T T I P	
Muk151QII28	GGAGCCGACGGCATTGCCTTCTTCATCGCACCGCCGGATACTACCATACC	323
101	S G S G G G L L G L F A P G T A	
Muk151QII28	GAGCGGTTTCGGGCGGAGGTCTCCTAGGGCTCTTGCACCTGGAAGTGTCTC	373
117	Q N T S A N Q V I A V E F D T F Y	
Muk151QII28	AAAACACGTCCGCGAACCAAGTGATTGCGGTTGAGTTTGACACCTTCTAT	423
134	A Q D S N T W D P N Y P H I G I D	
Muk151QII28	GCTCAAGACTCTAACACTTGGGATCCGAATTACCCGCACATTGGAATCGA	473
151	V N S I R S V K T V K W D R R D	
Muk151QII28	TGTCAACTCCATTAGGTCTGTGAAGACTGTGAAATGGGATAGGAGAGATG	523
167	G Q S L N V L V T F N P S T R N L	
Muk151QII28	GTCAATCCCTTAATGTCCTCGTAACATTTAACCCCTAGCACTCGAAACCTG	573
184	D V V A T Y S D G T R Y E V S Y E	
Muk151QII28	GATGTCGTTGCTACTTACTCTGATGGTACAAGATACGAGGTGTCTTATGA	623
201	V D V R S V L P E W V R V G F S	
Muk151QII28	GGTTGACGTGAGGTCCGTGCTTCCAGAATGGGTTAGAGTTGGATTCTCCG	673
217	G G V G N A A E <u>Y</u> Q T H T L E S W	
Muk151QII28	<u>GAGGCGTTGGCAACGCGGCCGAATATCAAACGCATACTCTTGAATCATGG</u>	723
	<i>Bst</i> ZI	
234	S F T S T L L Y T A	
Muk151QII28	TCTTTCACCTCAACCTTGTTGTACACTGCT TAG -----	756

Figure 4.21: Nucleotide and amino acid sequence of the mature mukwa seed lectin mutant carrying the UEA II specificity loop. The UEA II specificity loop amino acid residues are underlined and shown in bold, and USP9 is bordered. The first and last amino acids of the mature mukwa seed lectin, glutamine residues, have been replaced by methionine and a stop codon respectively.

1	M D S L S F G F P T F P S D Q K	
Muk151QII28	- ATG GATTCCCTTTTCCTTCGGCTTCCCTACGTTTCCTTCAGACCAGAAAA	73
17	N L I F Q G D A Q I K N N A V Q L	
Muk151QII28	ACCTAATCTTCCAAGGTGATGCACAGATTAAAAACAACGCAGTCCAGCTT	123
34	T K T D S N G N P V A S T V G R I	
Muk151QII28	ACCAAGACAGACAGTAATGGTAACCCAGTGGCAAGCACTGTTGGGCGAAT	173
51	L F S A Q V H L W E K S S S R V	
Muk151QII28	CTTATTCTCGGCGCAAGTGCACCTTTGGGAAAAAGCTCAAGCAGAGTGG	223
67	A N F Q S Q F S F S L K S P L S N	
Muk151QII28	CAAACCTTTCAATCCCAGTTCAGCTTTTCCCTCAAATCACCGCTTTCCAAT	273
84	G A D G I A F F I A P P D T T I P	
Muk151QII28	GGAGCCGACGGCATTGCCTTCTTCATCGCACCGCCGGATACTACCATACC	323
101	S G S G G G L L G L F A P G T A	
Muk151QII28	GAGCGGTTTCGGGCGGAGGTCTCCTAGGGCTCTTGCACCTGGAACCTGCTC	373
117	Q N T S A N Q V I A V E F D T F Y	
Muk151QII28	AAAACACGTCCGCGAACCAAGTGATTGCGGTTGAGTTTGACACCTTCTAT	423
134	A Q D S N T W D P N Y P H I G I D	
Muk151QII28	GCTCAAGACTCTAACACTTGGGATCCGAATTACCCGCACATTGGAATCGA	473
151	V N S I R S V K T V K W D R R D	
Muk151QII28	TGTCAACTCCATTAGGTCTGTGAAGACTGTGAAATGGGATAGGAGAGATG	523
167	G Q S L N V L V T F N P S T R N L	
Muk151QII28	GTCAATCCCTTAATGTCCTCGTAACATTTAACCTTAGCACTCGAAACCTG	573
184	D V V A T Y S D G T R Y E V S Y E	
Muk151QII28	GATGTCGTTGCTACTTACTCTGATGGTACAAGATACGAGGTGTCTTATGA	623
201	V D V R S V L P E W V R V G F S	
Muk151QII28	GGTTGACGTGAGGTCCGTGCTTCCAGAATGGGTTAGAGTTGGATTCTCCG	673
217	G A T G A Q R D A A Q T H T L E S	
Muk151QII28	<u>GAGCGACCGGTGCACAGCGTGATGCGGCTCAAACGCATACTCTTGAATCA</u> <i>ApaLI</i>	723
234	W S F T S T L L Y T A	
Muk151QII28	TGGTCTTTCACCTCAACCTTGTTGTACACTGCT TAG -----	759

Figure 4.22: Nucleotide and amino acid sequence of the mature mukwa seed lectin mutant carrying the ECorL specificity loop. The ECorL specificity loop amino acid residues are shown in bold and underlined. The first and last amino acids of the mature mukwa seed lectin, glutamine residues, have been replaced by methionine and a stop codon respectively.

1	M D S L S F G F P T F P S D Q K	
Muk151QII28	- ATG GATTCCCTTTCTTCGGCTTCCCTACGTTTCCTTCAGACCAGAAAA	73
17	N L I F Q G D A Q I K N N A V Q L	
Muk151QII28	ACCTAATCTTCCAAGGTGATGCACAGATTAAAAACAACGCAGTCCAGCTT	123
34	T K T D S N G N P V A S T V G R I	
Muk151QII28	ACCAAGACAGACAGTAATGGTAACCCAGTGGCAAGCACTGTTGGGCGAAT	173
51	L F S A Q V H L W E K S S S R V	
Muk151QII28	CTTATTCTCGGCGCAAGTGCACCTTTGGGAAAAAGCTCAAGCAGAGTGG	223
67	A N F Q S Q F S F S L K S P L S N	
Muk151QII28	CAAACCTTTCAATCCCAGTTCAGCTTTTCCCTCAAATCACCGCTTTCCAAT	273
84	G A D G I A F F I A P P D T T I P	
Muk151QII28	GGAGCCGACGGCATTGCCTTCTTCATCGCACCGCCGGATACTACCATACC	323
101	S G S G G G L L G L F A P G T A	
Muk151QII28	GAGCGGTTCTGGGCGGAGGTCTCTAGGGCTCTTTGCACCTGGAACCTGCTC	373
117	Q N T S A N Q V I A V E F D T F Y	
Muk151QII28	AAAACACGTCCGCGAACCAAGTGATTGCGGTTGAGTTTGACACCTTCTAT	423
134	A Q D Y N T W D P N Y P H I G I D	
Muk151QII28	GCGCAAGACT <u>TAT</u> AACACTTGGGATCCGAATTACCCGCACATTGGAATCGA	473
	<i>FspI</i>	
151	V N S I R S V K T V K W D R R D	
Muk151QII28	TGTCAACTCCATTAGGTCTGTGAAGACTGTGAAATGGGATAGGAGAGATG	523
167	G Q S L N V L V T F N P S T R N L	
Muk151QII28	GTCAATCCCTTAATGTCCTCGTAACATTTAACCTTAGCACTCGAAACCTG	573
184	D V V A T Y S D G T R Y E V S Y E	
Muk151QII28	GATGTCGTTGCTACTTACTCTGATGGTACAAGATACGAGGTGTCTTATGA	623
201	V D V R S V L P E W V R V G F S	
Muk151QII28	GGTTGACGTGAGGTCCGTGCTTCCAGAATGGGTTAGAGTTGGATTCTCAG	673
217	A A S G G Y Q T H T L E S W S F T	
Muk151QII28	CTGCCTCAGGAG <u>GGA</u> TATCAAACGCATACTCTTGAATCATGGTCTTTCACC	723
234	S T L L Y T A	
Muk151QII28	TCAACCTTGTTGTACACTGCT TAG -----	747

Figure 4.23: Nucleotide and amino acid sequence of the mature mukwa seed lectin mutant carrying the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution (shown in bold and underlined). The first and last amino acids of the mature mukwa seed lectin, glutamine residues, have been replaced by methionine and a stop codon respectively.

1	M D S L S F G F P T F P S D Q K	
Muk151QII28	- ATG GATTCCCTTTTCCTTCGGCTTCCCTACGTTTCCTTCAGACCAGAAAA	73
17	N L I F Q G D A Q I K N N A V Q L	
Muk151QII28	ACCTAATCTTCCAAGGTGATGCACAGATTAAAAACAACGCAGTCCAGCTT	123
34	T K T D S N G N P V A S T V G R I	
Muk151QII28	ACCAAGACAGACAGTAATGGTAACCCAGTGGCAAGCACTGTTGGGCGAAT	173
51	L F S A Q V H L W E K S S S R V	
Muk151QII28	CTTATTCTCGGCGCAAGTGCACCTTTGGGAAAAAAGCTCAAGCAGAGTGG	223
67	A N F Q S Q F S F S L K S P L S N	
Muk151QII28	CAAACCTTTCAATCCCAGTTCAGCTTTTCCCTCAAATCACCGCTTTCCAAT	273
84	G A D G I A F F I A P P D T T I P	
Muk151QII28	GGAGCCGACGGCATTGCCTTCTTCATCGCACCGCCGGATACTACCATAACC	323
101	S G S G G G L L G L F A P G T A	
Muk151QII28	GAGCGGTTTCGGGCGGAGGTCTCCTAGGGCTCTTTGCACCTGGAACCTGCTC	373
117	Q N T S A N Q V I A V E F D T F Y	
Muk151QII28	AAAACACGTCCGCGAACCAAGTGATTGCGGTTGAGTTTGACACCTTCTAT	423
134	A Q D Y N T W D P N Y P H I G I D	
Muk151QII28	GCGCAAGACT <u>TAT</u> AACACTTGGGATCCGAATTACCCGCACATTGGAATCGA	473
151	V N S I R S V K T V K W D R R D	
Muk151QII28	TGTCAACTCCATTAGGTCTGTGAAGACTGTGAAATGGGATAGGAGAGATG	523
167	G Q S L N V L V T F N P S T R N L	
Muk151QII28	GTCAATCCCTTAATGTCCTCGTAACATTTAACCCCTAGCACTCGAAACCTG	573
184	D V V A T Y S D G T R Y E V S Y E	
Muk151QII28	GATGTCGTTGCTACTTACTCTGATGGTACAAGATACGAGGTGTCTTATGA	623
201	V D V R S V L P E W V R V G F S	
Muk151QII28	GGTTGACGTGAGGTCCGTGCTTCCAGAATGGGTTAGAGTTGGATTCTCAG	673
217	A A S G <u>G</u> <u>G</u> Y Q T H T L E S W S F	
Muk151QII28	CTGCCTCAGGAG <u>GGAGGA</u> TATCAAACGCATACTCTTGAATCATGGTCTTTC	723
234	T S T L L Y T A	
Muk151QII28	ACCTCAACCTTGTTGTACACTGCT TAG -----	750

Figure 4.24: Nucleotide and amino acid sequence of the mature mukwa seed lectin mutant carrying the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution (shown in bold and underlined). The first and last amino acids of the mature mukwa seed lectin, glutamine residues, have been replaced by methionine and a stop codon respectively.

4.5 Analysis of Expression of the Mukwa Seed Lectin Mutants in *E. coli*

For expression, the pBADMycHisA-mukwa seed lectin mutants pGV5031, pGV5022, pGV5018 and pGV5026 representing the UEA II specificity loop, ECorL specificity loop, $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution and $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution respectively were separately transformed into *E. coli* TOP10 cells. Varying amounts of L-arabinose from 0.00002% to 0.2% (w/v) were then used to determine the appropriate amount of arabinose required for maximum expression of the mukwa seed lectin mutants.

Analysis of expression was carried out using SDS-PAGE. The SDS-PAGE gave the expected band of approximately 25 kDa in size for all the four mutants. This band increased in intensity with increasing concentration of L-arabinose as can be seen in each of Figures 4.25, 4.26, 4.27 and 4.28. Figure 4.25 shows that the UEA II specificity loop mutant gave the weakest band suggesting that expression of this mutant gave the least protein. The intensity of the band pattern corresponding to the expressed protein in Figure 4.26 shows that the ECorL specificity loop mutant is expressed in the highest quantities.

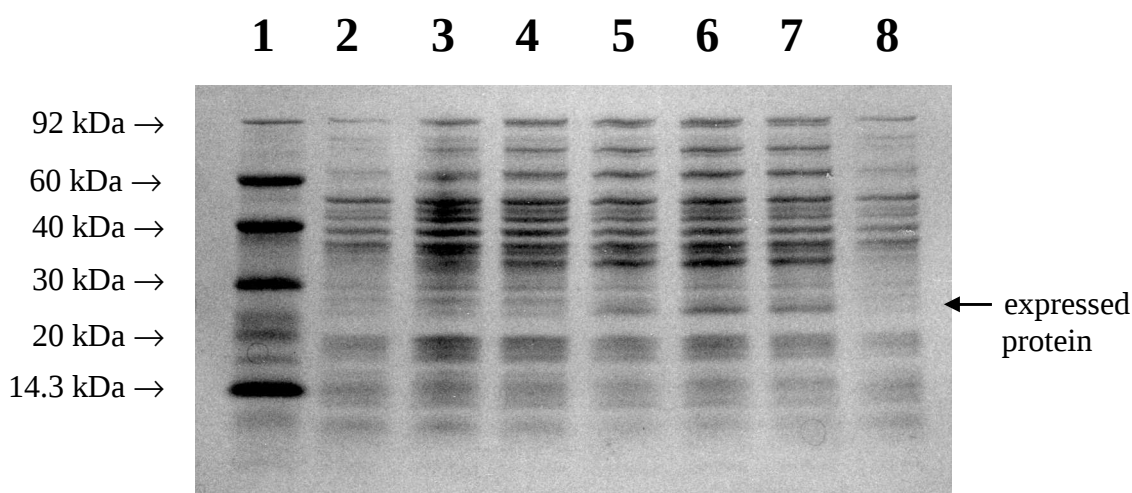


Figure 4.25: SDS-PAGE analysis of total protein in *E. coli* TOP10 cells transformed with pGV5031, a pBADMycHisA clone carrying the cDNA encoding the mukwa seed lectin mutant with the UEA II specificity loop. Lane 1 shows the protein MM size markers. Lanes 2 and 8 show the total protein before induction of expression. Lanes 3 to 7 show the total protein obtained after induction with ten-fold increment of L-arabinose, from 0.00002% to 0.2% (w/v).

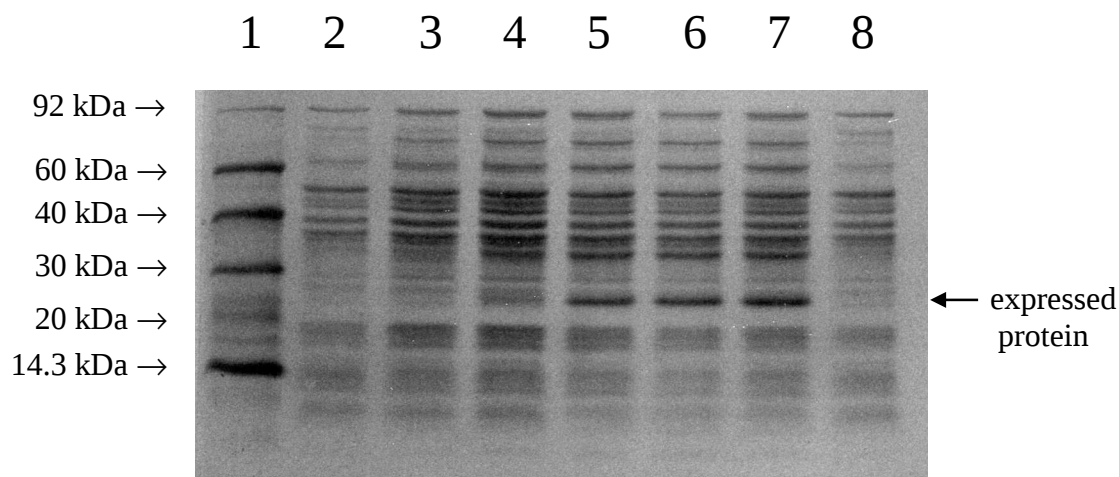


Figure 4.26: SDS-PAGE analysis of total protein in *E. coli* TOP10 cells transformed with pGV5022, a pBADMycHisA clone carrying the cDNA encoding the mukwa seed lectin mutant with the ECorL specificity loop. Lane 1 shows the protein MM size markers. Lanes 2 and 8 show the total protein before induction of expression. Lanes 3 to 7 show the total protein obtained after induction with ten-fold increment of L-arabinose, from 0.00002% to 0.2% (w/v).

Comparison of Figures 4.27 and 4.28 shows that the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G mutant was expressed in higher quantities than the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G mutant.

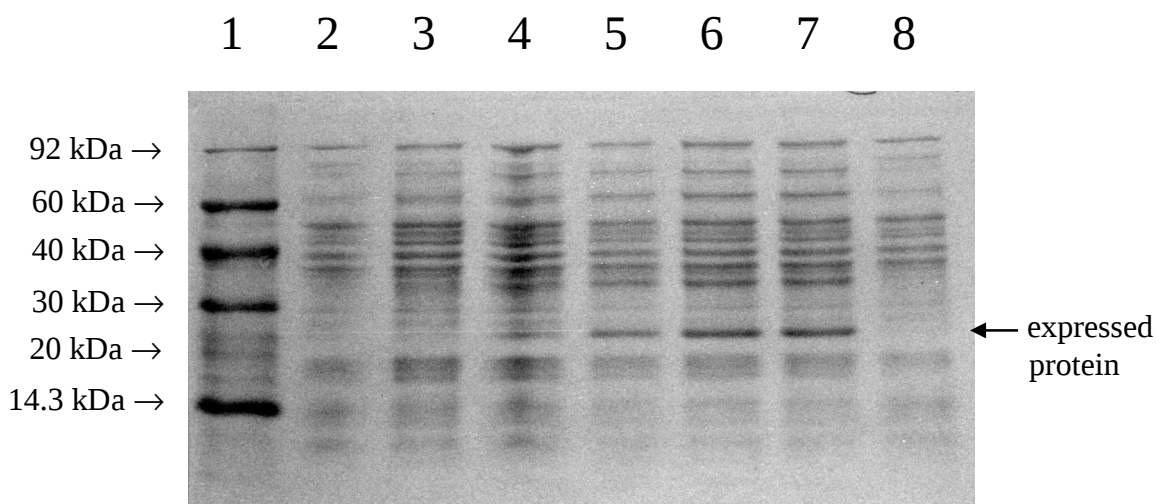


Figure 4.27: SDS-PAGE analysis of total protein in *E. coli* TOP10 cells transformed with pGV5018, a pBADMycHisA clone carrying the cDNA encoding the mukwa seed lectin mutant with the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution in the specificity loop. Lane 1 shows the protein MM size markers. Lanes 2 and 8 show the total protein before induction of expression. Lanes 3 to 7 show the total protein obtained after induction with ten-fold increment of L-arabinose, from 0.00002% to 0.2% (w/v).

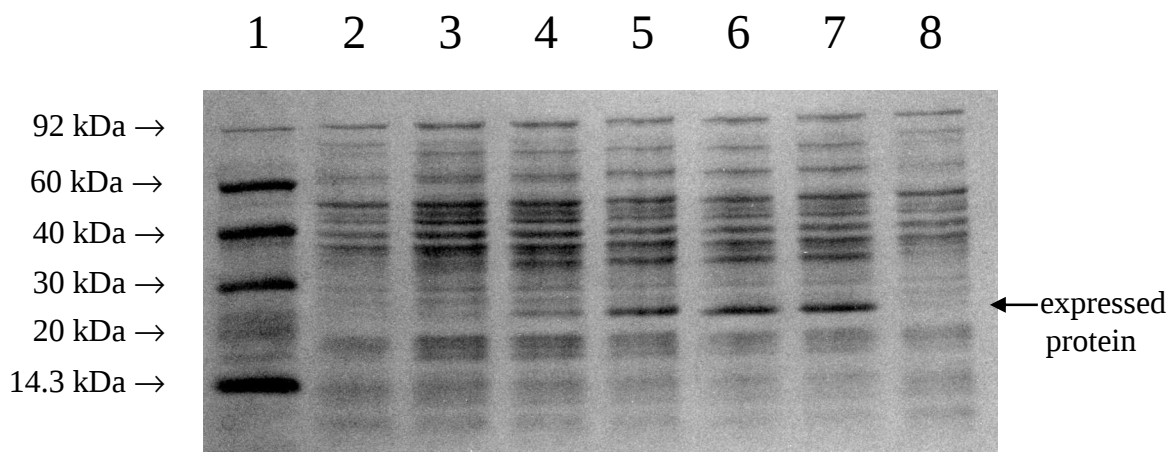


Figure 4.28: SDS-PAGE analysis of total protein in *E. coli* TOP10 cells transformed with pGV5026, a pBADMyHisA clone carrying the cDNA encoding the mukwa seed lectin mutant with the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution in the specificity loop. Lane 1 shows the protein MM size marker. Lanes 2 and 8 show the total protein before induction of expression. Lanes 3 to 7 show the total protein obtained after addition of ten-fold increment of L-arabinose, from 0.00002% to 0.2% (w/v).

The control mutants, pGV5016, pGV5021 and pGV5030 representing the $S^{137} \rightarrow Y$ change only, $E^{221}Q^{222} \rightarrow$ single G change only and $E^{221}Q^{222} \rightarrow$ double G change only respectively, also gave the same band pattern upon expression in *E. coli* TOP10 cells (data not shown).

4.6 Characterization of the Mukwa Seed Mutant Lectins

4.6.1 Cross-Reactivity of the Mukwa Seed Recombinant Lectins with Antiserum Directed Against the Native Mukwa Seed Lectin

The pBADMyHisA-mukwa seed lectin recombinants pGV4977, pGV5031, pGV5022, pGV5018 and pGV5026 representing the wild type, the UEA II specificity loop, ECorL specificity loop, $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution and $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution respectively were separately transformed into *E. coli* TOP10 cells. The crude extract from *E. coli* cells expressing each of the mukwa seed recombinant lectins was subjected

to 0-30%, 30-60% and 60-90% ammonium sulphate precipitation, and the different fractions used in agarose gel immunodiffusion tests. Rabbit antiserum directed against the mukwa seed lectin was used as the source of specific antibodies. Pre-immune serum and the native mukwa seed lectin were included as negative and positive controls respectively.

For all the recombinant lectins except the UEA II specificity loop mutant, strong precipitin lines formed with the 30-60% fractions while faint ones could be observed with the 60-90% fractions (see Figure 4.29). The recombinant lectins therefore precipitate mainly in the 30-60% ammonium sulphate range, just like the native seed lectin.

The immunodiffusion tests showed that there is a cross-reaction between the mukwa seed recombinant lectins and antibodies directed against the native lectin isolated from mukwa seeds. The UEA II specificity loop mutant was, however, probably expressed in too low a quantity to be detected by immunodiffusion tests. The recombinant lectins could therefore be purified by affinity chromatography on an antibody column.

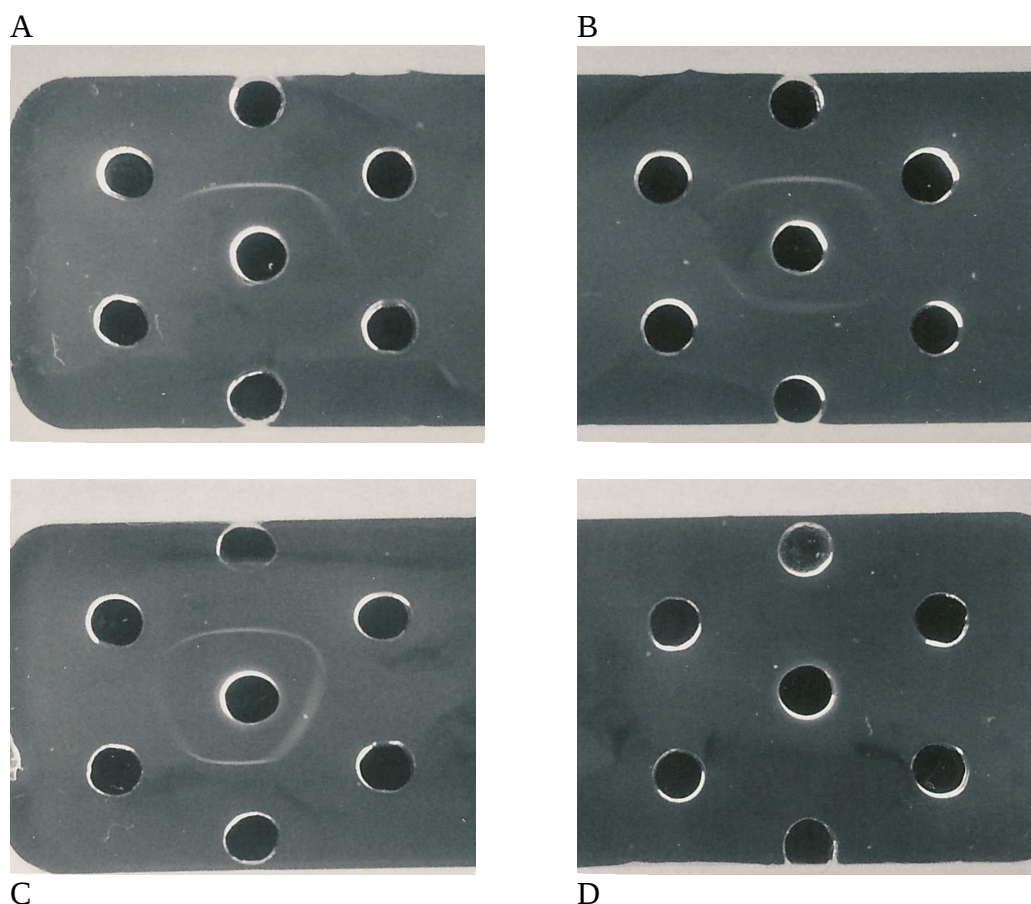
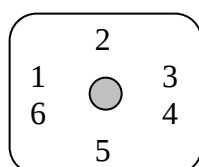


Figure 4.29: Immunodiffusion assays of the ammonium sulphate (AS) fractions of the crude extract from *E. coli* TOP10 cells expressing the mukwa seed recombinant lectins with rabbit anti-native mukwa seed lectin serum. The central hole in A, B and C contained antiserum directed against the mukwa seed lectin while in D it contained pre-immune serum.

Rosette



In A, holes 1-3 contained the wild type AS fractions 0-30%, 30-60% and 60-90% respectively. Corresponding AS fractions of other recombinant lectins were placed as follows: the UEA II specificity loop mutant in A, holes 4-6, the ECorL specificity loop mutant in B, holes 1-3, the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution mutant in B, holes 4-6 and the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution mutant in C, holes 1-3 and in D, holes 1-3. Holes 4-6 in C and in D contained the purified native mukwa seed lectin at concentrations 1 mg/ml, 0.2 mg/ml and 0.04 mg/ml respectively.

4.6.2 Preparation of Glycoproteins of High Variability

Diethylaminoethyl-cellulose was used to separate different fractions of glycoproteins from porcine and bovine plasma. Porcine plasma gave a slightly different elution profile from bovine plasma. For both porcine and bovine plasma, the unbound fraction, designated Peak 1 was eluted with the void volume while the BSA fraction, designated Peak 2, was eluted with 25 mM Na-acetate buffer, pH 4.4. For porcine plasma, a small peak, designated Peak 3 was obtained just before adding the end buffer, Na-acetate, pH 4.0. Elution with the end buffer gave a fraction which was designated Peak 4 for both porcine and bovine plasma.

4.6.3 Recognition of Glycoproteins by the Mukwa Seed Recombinant Lectins

Different glycoproteins were blotted onto nitrocellulose, incubated with the ammonium sulphate fractions containing the recombinant lectins, and immunoprobed with antiserum directed against the native mukwa seed lectin. This assay was carried out to assess if the mukwa seed mutant lectins were capable of binding any glycoproteins.

All the recombinant lectins were capable of binding glycoproteins as can be seen in Figures 4.30 to 4.34. The different band patterns observed in these figures suggest that the recombinant lectins recognize different sugar moieties on glycoproteins. Bands of the highest intensity can be seen in Panel A of Figure 4.30 indicating that the wild type mukwa seed lectin has the highest affinity for the glycoproteins that were assayed. As this wild type lectin is mannose/glucose specific, the bands observed on the blot processed in the presence of methyl- α -D-mannopyranoside (see Figure 4.30 Panel B) have a higher affinity for the protein than for the sugar mannose. Panel C of Figure 4.30 and Panels B of Figures 4.31 to 4.34 show a non-specifically stained band, which was persistent even when the sugar moieties had been destroyed by pre-treatment with NaIO₄. It follows that this band, which is observed in all the other panels (shown by the arrow), is a result of protein-protein (lectin-protein) interaction and not protein-glycan (lectin-sugar) interaction.

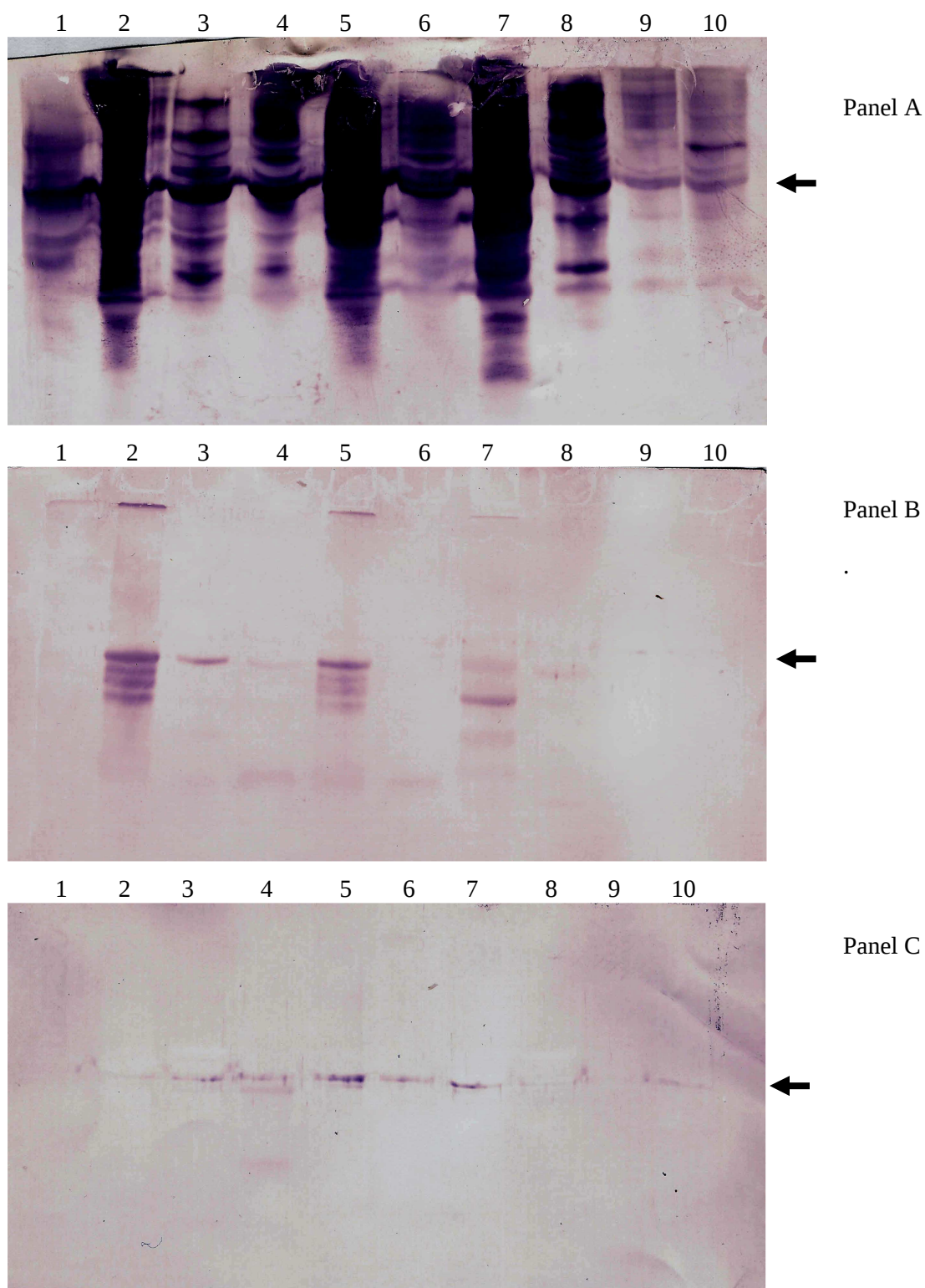


Figure 4.30: Western blot analysis of the recognition of glycoproteins by the 30-60% ammonium sulphate fraction of the crude extract from *E. coli* TOP 10 cells expressing the wild type mukwa seed lectin. Lane 1 shows hen egg white proteins, lanes 2-6 show the bands from the respective porcine plasma IgG fraction, peak 1, peak 2, peak 3 and peak 4. Lanes 7-10 show the bands from the respective bovine plasma IgG fraction, peak 1, peak 2 and peak 4. After incubation with the lectin sample, immunoprobining was done using rabbit antiserum directed against the native mukwa seed lectin followed by anti-rabbit IgG phosphatase. Panel A shows the blot processed normally. Panel B shows the blot processed in the presence of 100 mM methyl- α -D-mannopyranoside. Panel C shows the blot treated with NaIO_4 prior to incubation with the lectin sample and immunoprobining.

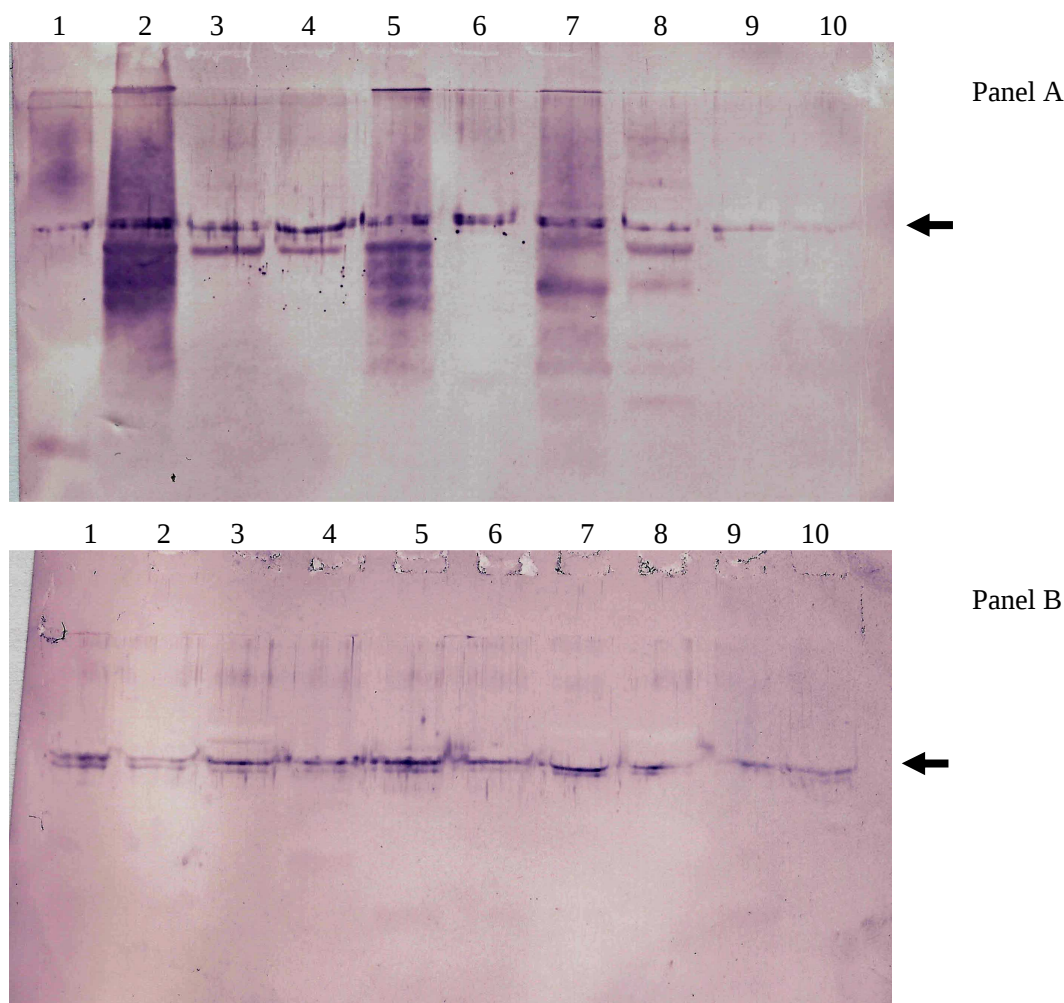


Figure 4.31: Western blot analysis of the recognition of glycoproteins by the 30-60% ammonium sulphate fraction of the crude extract from *E. coli* TOP 10 cells expressing the mukwa seed lectin mutant carrying the UEA II specificity loop. Lane 1 shows hen egg white proteins, lanes 2-6 show the bands from the respective porcine plasma IgG fraction, peak 1, peak 2, peak 3 and peak 4. Lanes 7-10 show the bands from the respective bovine plasma IgG fraction, peak 1, peak 2 and peak 4. After incubation with the lectin sample, immunoprobings were done using rabbit antiserum directed against the native mukwa seed lectin followed by anti-rabbit IgG phosphatase. Panel A shows the blot processed normally while Panel B shows the blot treated with NaIO_4 prior to incubation with the lectin sample and immunoprobings.

From Panel A of Figure 4.31, bands of high intensity can be observed mainly in lanes 2, 5 and 7 corresponding to the porcine plasma IgG and peak 3, and bovine plasma IgG respectively. The mukwa seed lectin mutant carrying the UEA II specificity loop therefore binds strongly to the sugar moieties of the glycoproteins found in these fractions.

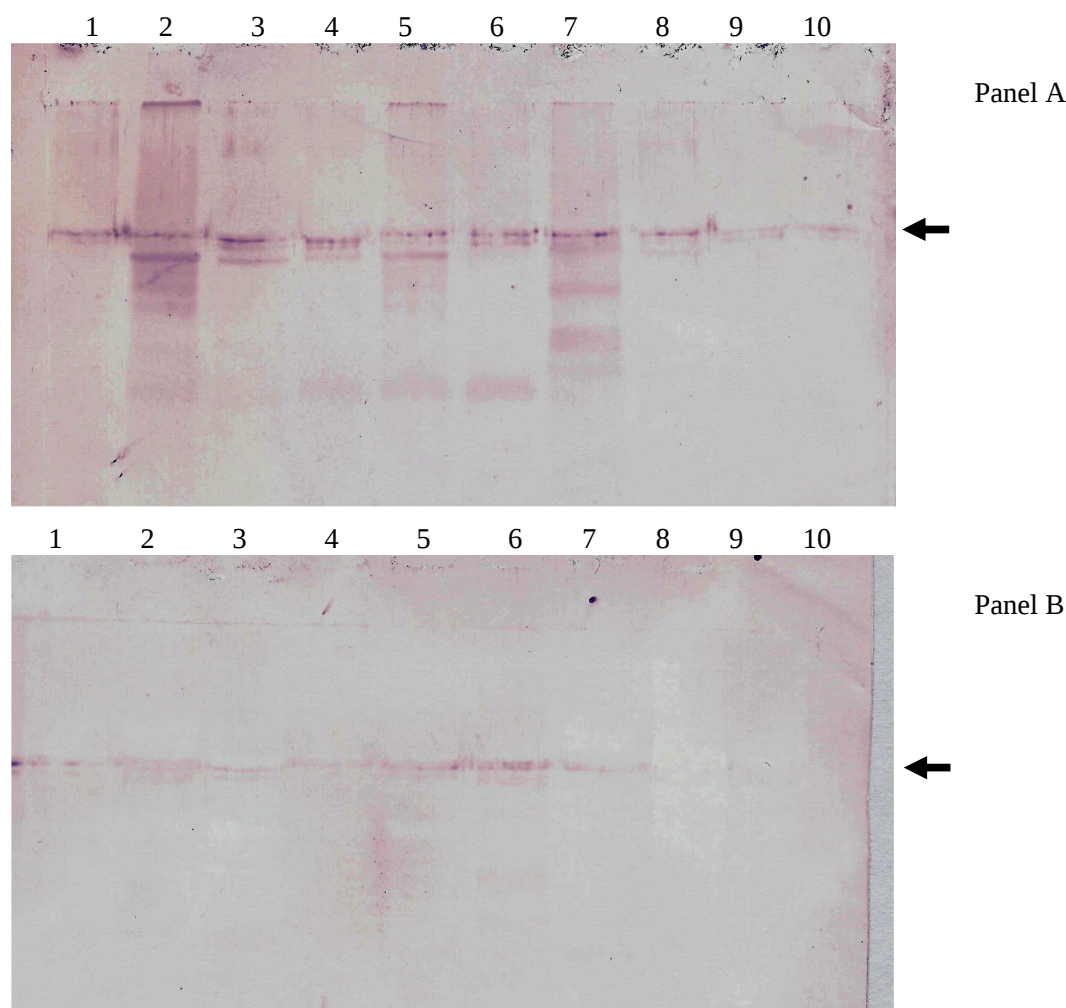


Figure 4.32: Western blot analysis of the recognition of glycoproteins by the 30-60% ammonium sulphate fraction of the crude extract from *E. coli* TOP 10 cells expressing the mukwa seed lectin mutant carrying the ECorL specificity loop. Lane 1 shows hen egg white proteins, lanes 2-6 show the bands from the respective porcine plasma IgG fraction, peak 1, peak 2, peak 3 and peak 4. Lanes 7-10 show the bands from the respective bovine plasma IgG fraction, peak 1, peak 2 and peak 4. After incubation with the lectin sample, immunoprobining was done using rabbit antiserum directed against the native mukwa seed lectin followed by anti-rabbit IgG phosphatase. Panel A shows the blot processed normally while Panel B shows the blot treated with NaIO₄ prior to incubation with the lectin sample and immunoprobining.

Of the four mutants, the mukwa seed lectin mutant carrying the ECorL specificity loop gave bands of the lowest intensity as can be seen in Panel A of Figure 4.32, suggesting that this mutant has the lowest affinity for the glycoproteins that were assayed.

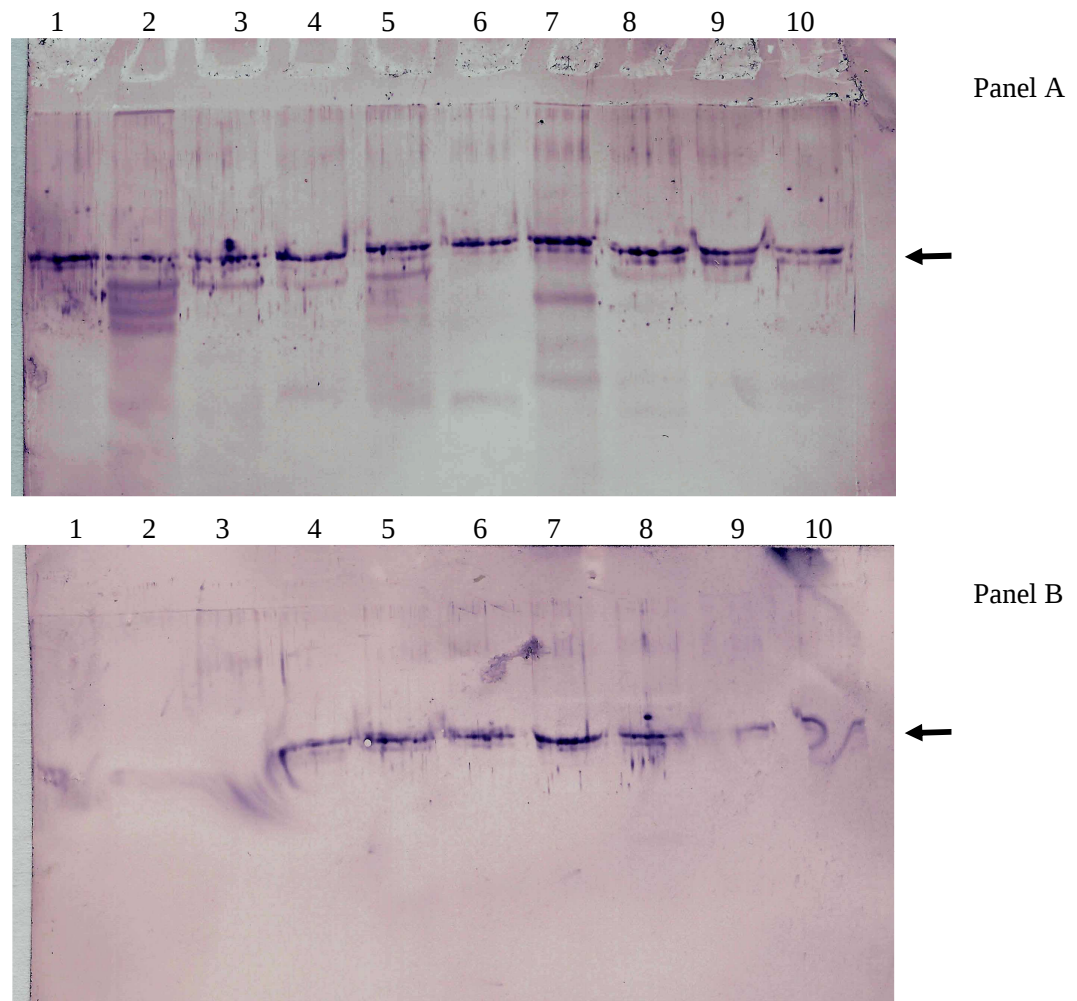


Figure 4.33: Western blot analysis of the recognition of glycoproteins by the 30-60% ammonium sulphate fraction of the crude extract from *E. coli* TOP 10 cells expressing the mukwa seed lectin mutant carrying the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution. Lane 1 shows hen egg white proteins, lanes 2-6 show the bands from the respective porcine plasma IgG fraction, peak 1, peak 2, peak 3 and peak 4. Lanes 7-10 show the bands from the respective bovine IgG fraction, peak 1, peak 2 and peak 4. After incubation with the lectin sample, immunoprobining was done using rabbit antiserum directed against the native mukwa seed lectin followed by anti-rabbit IgG phosphatase. Panel A shows the blot processed normally while Panel B shows the blot treated with $NaIO_4$ prior to incubation with the lectin sample and immunoprobining.

Comparison of Figures 4.33 and 4.34 Panel A, shows that bands of a higher intensity were obtained with the mukwa seed lectin mutant carrying $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution than those of the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution mutant. The band pattern shown in these figures also appears to be slightly different.

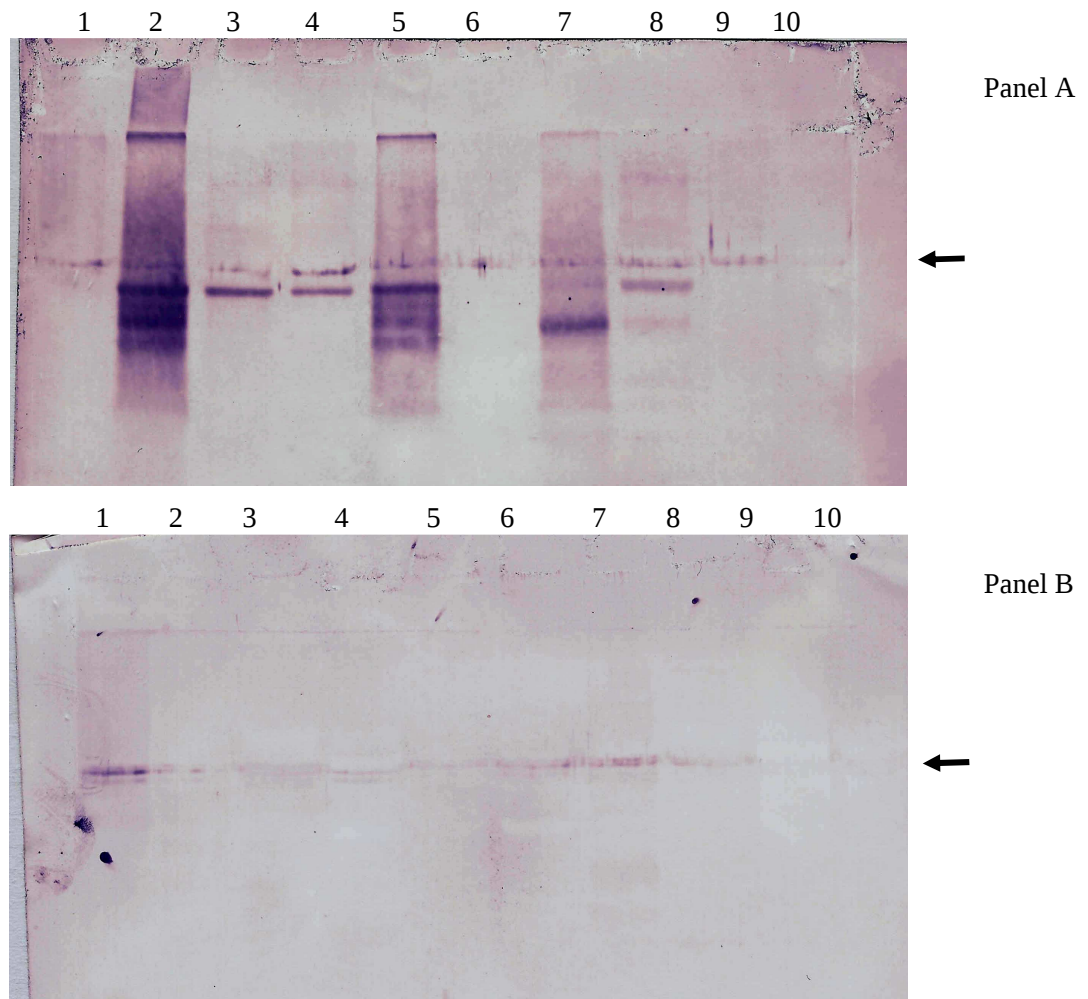


Figure 4.34: Western blot analysis of the recognition of glycoproteins by the 30-60% ammonium sulphate fraction of the crude extract from *E. coli* TOP 10 cells expressing the mukwa seed lectin mutant carrying the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution. Lane 1 shows hen egg white proteins, lanes 2-6 show the bands from the respective porcine plasma IgG fraction, peak 1, peak 2, peak 3 and peak 4. Lanes 7-10 show the bands from the respective bovine IgG fraction, peak 1, peak 2 and peak 4. After incubation with the lectin sample, immunoprobining was done using rabbit antiserum directed against the native mukwa seed lectin followed by anti-rabbit IgG phosphatase. Panel A shows the blot processed normally while Panel B shows the blot treated with $NaIO_4$ prior to incubation with the lectin sample and immunoprobining.

4.6.4 Preparation of Antiserum Directed Against the Native Mukwa Seed Lectin

Purification of the native mukwa lectin from a 30-60% ammonium sulphate precipitation fraction of the seed extract on Sephadex G-75 gave the expected elution profile (Manyumwa, 1998). The SDS-PAGE analysis and agglutination tests confirmed that the lectin was present in the fourth retarded peak.

Eight weeks after immunization, agarose gel immunodiffusion tests showed the presence of specific antibodies. Precipitin lines observed between the native mukwa seed lectin and the antiserum of each of the 3 rabbits were absent with the pre-immune sera.

4.6.5 Isolation of Specific Anti-Mukwa Seed Lectin IgG

Initially, caprylic acid fractionation was used to isolate total IgG from the bulk of serum proteins. Further purification of the caprylic acid fraction was achieved by ion-exchange chromatography at pH 6.2 on DEAE-cellulose. The positively charged IgG was eluted with the void volume and designated Peak 1, while the bound contaminants, eluted with the running buffer containing 2 M NaCl were designated Peak 2.

Specific IgG was then isolated from Peak 1 by affinity chromatography on a mukwa seed lectin-Sepharose-4B column. About 28 mg of the lectin were successfully coupled to 3 g of CNBr-activated Sepharose-4B. The specific IgG constituted only about 5% of the total protein in Peak 1.

4.6.6 Purification of the Mukwa Seed Recombinant Lectins on a Specific Anti-Mukwa Seed Lectin IgG-Sepharose-4B Column

Approximately 15 mg of specific IgG directed against the native mukwa seed lectin were successfully coupled onto 4 g of CNBr-activated Sepharose-4B. Purification of the recombinant lectins from a crude *E. coli* extract by affinity chromatography on this column gave different yields. From a 10 x 50 ml culture, the wild type lectin gave the highest yield of about 1.8 mg followed by the ECorL specificity loop mutant with a yield of about 1.6 mg. Both the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution and the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution mutants gave a yield of about 1 mg protein respectively. The UEA II specificity loop mutant gave hardly detectable yields.

Figures 4.35 and 4.36 show the analysis, by SDS-PAGE, of the purification process of the recombinant lectins.

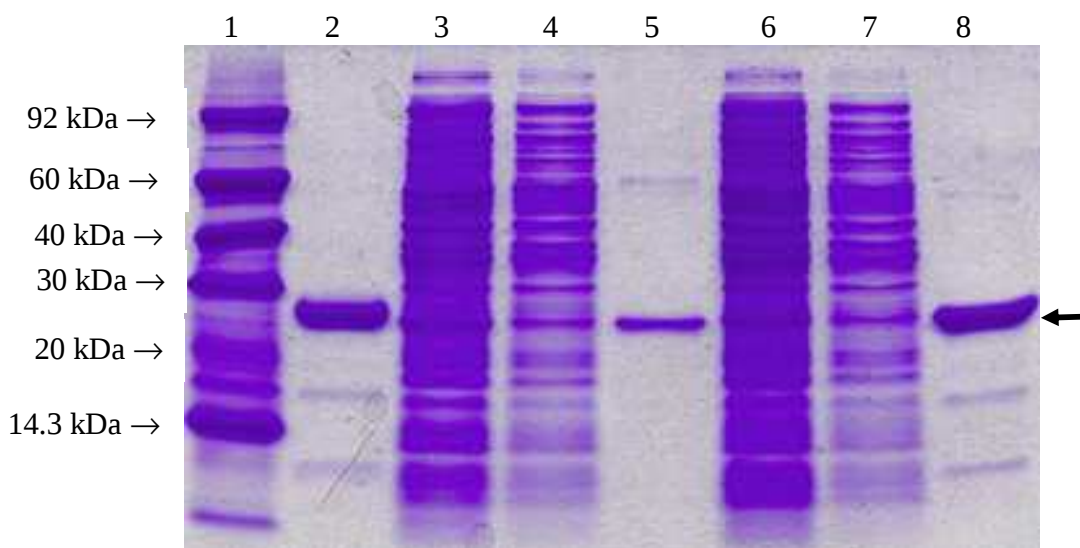


Figure 4.35: SDS-PAGE analysis of the purified wild type (lanes 3-5) and the ECorL specificity loop (lanes 6-8) mukwa seed recombinant lectins expressed in *E. coli* TOP10 cells. Purification was by affinity chromatography on a specific anti-mukwa seed lectin IgG-Sepharose-4B column. Lane 1 shows the protein MM size markers. Lane 2, the positive control, shows the native lectin isolated from mukwa seeds. Lanes 3 and 6 show the crude extract from *E. coli* TOP10 cells expressing the recombinant lectins. Lanes 4 and 7 show the unbound fractions. Lanes 5 and 8 show the lectin fraction eluted with 0.1 M acetic acid.

The expressed lectin band, approximately 25 kDa in size, can be seen in every sample that was loaded. The fact that this band can be observed even in the lanes corresponding to the unbound fractions indicates that the column was over-loaded. Given that the same control sample was loaded in lanes 2 of both Figures 4.35 and 4.36, it can be concluded that the dark staining seen at the top part of the gel shown in Figure 4.36 is probably aggregated material and is non-specific and anomalous.

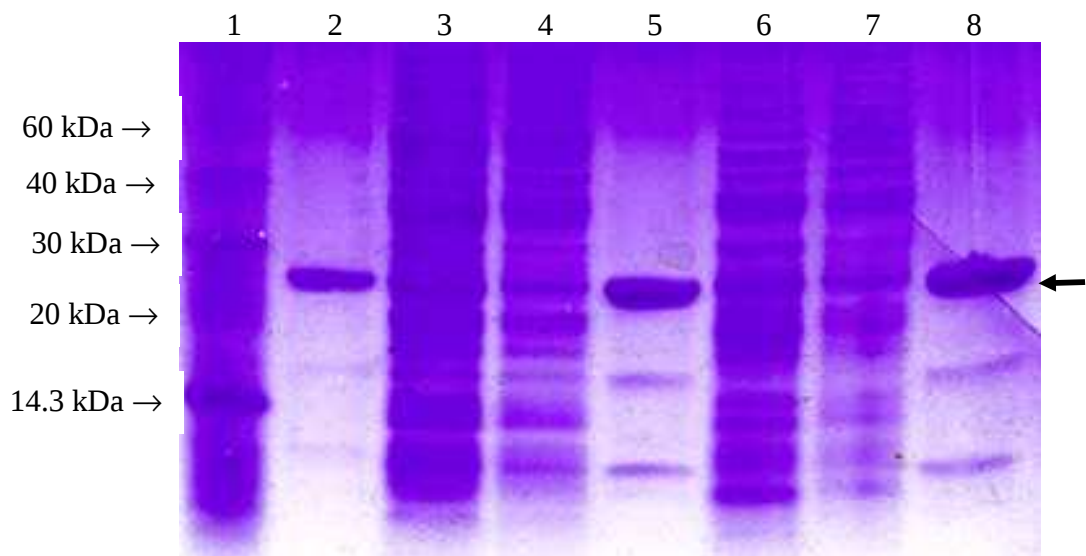


Figure 4.36: SDS-PAGE analysis of the purified $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution (lanes 3-5) and the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution (lanes 6-8) mukwa seed mutant lectins expressed in *E. coli* TOP10 cells. Purification was by affinity chromatography on a specific anti-mukwa seed lectin IgG-Sepharose-4B column. Lane 1 shows the protein MM size markers. Lane 2, the positive control, shows the native lectin isolated from mukwa seeds. Lanes 3 and 6 show the crude extract from *E. coli* TOP10 cells expressing the mutant lectins. Lanes 4 and 7 show the unbound fractions. Lanes 5 and 8 show the lectin fraction eluted with 0.1 M acetic acid.

4.6.7 Agglutinating Assays of the Mukwa Seed Mutant Lectins

All the three purified mutant lectins, namely the ECorL specificity loop, the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution and the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution mutants did not agglutinate any of the different animal erythrocytes tested. The native mukwa seed lectin and the wild type recombinant lectins, included as controls, agglutinated rabbit and sheep erythrocytes but not cow, goat or pig red blood cells.

The 30-60% ammonium sulphate fractions of the crude extract from *E. coli* cells expressing the wild type and the mutant lectins did not agglutinate any of the untreated human erythrocytes of the blood groups A1+, A2+, B+, AB+ and O+.

5. DISCUSSION

5.1 Expression of the cDNA Encoding the Mukwa Seed Lectin in *E. coli*

Mukwa seeds contain large amounts of the lectin such that a yield of 4 mg per g of defatted seed meal can be obtained by affinity chromatography on mannose-Sepharose-4B. Efforts to express the seed lectin were made in order to develop an experimental system in which specific amino acid alterations could be introduced into the sugar-binding site. Milligram quantities of the recombinant lectins would be required for their three-dimensional structures to be determined by X-ray crystallography.

Bacteria, particularly *E. coli*, which has been well-characterized, are the first choice of an expression system when one desires to produce a given protein in large quantities. The *E. coli* grows at a very fast rate in comparison with other expression systems such as mammalian cells, giving the opportunity to purify, analyze and use the expressed protein in a much shorter time. Transformation of *E. coli* cells with foreign DNA is easy, requires minimal amounts of DNA and is generally inexpensive (Verma *et al.*, 1998; Sambrook and Russell, 2001).

Expression of eukaryotic genes in bacteria, however, has its own drawbacks. Most legume lectins are glycosylated and as bacteria are not capable of carrying out post-translational modifications, the expressed protein may not be functional. Production of the protein may be toxic to the bacteria leading to loss or mutation of the gene or cDNA. Inducible promoters such as the *lac* promoter are therefore normally used to control expression of the protein (Verma *et al.*, 1998). In spite of the controlled expression, at times the protein will precipitate in inclusion bodies within the bacteria and must be solubilized and re-folded (Lilie *et al.*, 1998). For instance, production of the pea lectin in *E. coli* resulted in the formation of insoluble aggregates in the host and the functional

lectin was recovered by solubilization of the aggregates in guanidium hydrochloride and renaturation in the presence of MnCl_2 and CaCl_2 (Stubbs *et al.*, 1986). Solubilization and renaturation are, however, sometimes only partially successful (Lilie *et al.*, 1998). In such cases, yeast is chosen as the ideal expression system as this host is capable of glycosylating proteins as well as providing an advanced protein-folding pathway for heterologous proteins. The yeast *Pichia pastoris* has been used to produce gram quantities of recombinant *P. vulgaris* phytohemagglutinin E-form (Baumgartner *et al.*, 2002).

In spite of the limitations that are associated with the expression of plant proteins in bacterial systems, some legume lectins such as *E. corallodendron* lectin (Arango *et al.*, 1992) and the lima bean lectin (Jordan and Goldstein, 1994) have however, been successfully expressed in *E. coli*. These expressed lectins are devoid of their glycosyl moieties yet they are biologically active. We therefore set out to express the cDNA of the mature mukwa seed lectin in *E. coli*.

The cDNA encoding part of the signal peptide, the mature mukwa seed lectin, the C-terminal peptide and the 3' untranslated region was available as a clone designated Muk151QII28. Clone Muk151QII28 was used as template to amplify, only the DNA region that codes for the mature mukwa seed lectin, by PCR using each of primers Muk-23 and Muk-24 with Muk-25 respectively. *Ex TaqTM* DNA polymerase was used for PCR as it has proof-reading activity and the recommended lower temperature of 68°C for elongation allows this step to take place slowly (Takara Biomedicals, 2000). High fidelity was important so as to avoid the incorporation of undesirable mutations. Muk-25 included a stop codon and an *EcoRI* restriction endonuclease recognition site. The polymerase chain reaction using Muk-23 and Muk-25 allowed for amplification of the coding sequence from Q¹ to A²⁴¹, the first to the second last amino acid residues of the mature lectin. The polymerase chain reaction using Muk-24 and Muk-25 allowed for amplification of the

coding sequence from the aspartic acid, the second amino acid residue, D², to A²⁴¹ the second last residue, thereby leaving out the nucleotide sequence coding for the first and last glutamine residues. The last amino acid residue of the mature lectin, Q²⁴², could not be identified from the crystal structure. Both amplification products excluded the nucleotide sequence coding for the C-terminal peptide, residues K²⁴³ to M²⁵², as this could not be seen in X-ray crystallography (Loris *et al.*, 2003; 2004). It could therefore be concluded that the C-terminal peptide is either too flexible to be detected using this technique (Neumann *et al.*, 2004) or is proteolytically cleaved off to form the mature protein. Many legume lectins are proteolytically processed at their C-terminus, often leading to heterogeneity at this end of the protein (Van Driessche, 1988; Young *et al.*, 1995).

In designing the primers used in PCR, we used information obtained from X-ray crystals of the native mukwa seed lectin. The three-dimensional structure had shown that the N-terminal amino acid residue of the mature lectin is in the form of a cyclic glutamine, pyroglutamate, located in a pocket in the interior of the protein. This was confirmed by the fact that the N terminus of the affinity-purified protein was blocked (Loris *et al.*, 2004). It was assumed that the space in the pocket would likely be insufficient to accommodate the methionine encoded by the initiator codon from the expression vector. As methionine is required to initiate protein synthesis in bacteria (Laursen *et al.*, 2005), it was envisaged that leaving out the first amino acid of the mature lectin would probably create space for the methionine. However, as all this was based on speculation, we decided to design two 5' end primers, Muk-23 and Muk-24. This meant that if both fragments were cloned, two proteins would be expressed. The expectation, however, was that one of the proteins would possibly fold properly whereas the other would not have the correct conformation for folding.

The amplified DNA was then ligated with the expression vector pBADMycHisA. This vector carries the *araBAD* promoter (P_{BAD}) that is turned on by L-arabinose and therefore provides a tight, dose-dependent regulation of heterologous gene expression (Guzman *et al.*, 1995). For expression, the pBADMycHisA vector was digested with *NcoI* and filled in with T4 DNA polymerase in order to provide an ATG codon for methionine preceeding the cDNA coding for the mature lectin. Digestion of the expression vector and the amplification products with *EcoRI* made directional cloning possible. Inclusion of a stop codon in primer Muk-25 allowed for the termination of translation of the mature lectin, although it left out the last amino acid residue, glutamine. Termination of translation meant that the lectin could be expressed without the C-terminal peptide of the expression vector.

Clones were sequenced to check if the reading frame of the expression vector, pBADMycHisA, had been maintained at the point of ligation. The nucleotide sequence showed that the reading frame was preserved only in constructs such as pGV4976 and pGV4977 in which the codon for the first amino acid residue had been omitted. Constructs in which the insert was generated using PCR primers Muk-23 and Muk-25, that is, in which the glutamine residue was next to the initiating methionine were not recovered most probably due to suspected lethal effects to the host cells. Experience in the laboratory has shown that at times bacterial cells fail to grow if they carry certain constructs (H. De Greve, personal communication). For such cells to be viable, either a mutation occurs or the orientation of the construct changes. In this case, a particular orientation was forced due to directional cloning and hence only those cells harbouring the mutated construct survived. The mutation, however, meant that these constructs were out of translational frame. The construct pGV4977 that was subsequently used for the expression experiments had the nucleotide sequence coding for the first and last amino acid residues of the mature

mukwa seed lectin missing. In spite of the missing amino acid residues, the hope was that this wild type recombinant lectin would still be able to fold correctly.

For expression, pBADMyHis-wild type mukwa seed lectin recombinants were used to transform the *recA*, *endA* *E. coli* TOP10 strain. These host cells are capable of transporting L-arabinose but do not metabolize it and hence a constant level of arabinose can be maintained inside the host cells. The expression levels of the lectin could therefore be optimized to ensure maximum expression of soluble protein by varying the concentration of L-arabinose (Invitrogen Life Technologies, 2000). Maximum expression of the lectin, as shown by SDS-PAGE and by Western blotting, was obtained using 0.2% (w/v) L-arabinose and growing the culture overnight, for approximately 16 hours.

Analysis of the coding sequence of the mature mukwa seed lectin had shown the presence of seven arginine codons AGA/AGG. These arginine codons are rarely used in *E. coli* but their expression can be enhanced by co-transforming host cells with a second plasmid BL21-CodonPlus-RIL (Stratagene, 2000). Double transformants carrying both the pBADMyHisA-wild type mukwa seed lectin recombinant, pGV4977 and plasmid BL21-CodonPlus-RIL were therefore also expressed to see if levels of the mukwa seed lectin production could be increased. Analysis of expression on SDS-PAGE, however, showed that the addition of plasmid BL21-CodonPlus-RIL did not significantly enhance expression of the cDNA encoding the mukwa seed lectin. On the contrary, the presence of BL21-CodonPlus-RIL resulted in a very pronounced *E. coli* protein band, approximately 3 kDa smaller than the protein of interest. Thereafter, the pBADMyHisA-wild type mukwa seed lectin recombinants were expressed alone.

The expressed lectin agglutinated rabbit erythrocytes and the agglutination was inhibited by methyl- α -D-mannopyranoside. This biological test confirmed that the recombinant lectin was just as active as the native lectin from mukwa seeds. As *E. coli*

normally produces proteases in its cell envelope that may degrade the expressed protein after the cells are lysed (Gottesman, 1996), a separate experiment in which a cocktail of protease inhibitors was added to the bacterial extract was set up. The presence of protease inhibitors in the extract did not increase the yield of the expressed lectin and so thereafter the extract was prepared in their absence.

The exact yield of the expressed lectin in milligrams corresponded with the number of agglutinating units shown but varied from experiment to experiment. The slight differences in yield may have been due to the fact that induction of expression was carried out over a range 0.40 to 0.45 of optical density at 660 nm. Expression therefore may have been induced at slightly different points in the bacterial mid-log phase of the growth curve.

Higher yields of the expressed lectin were obtained from a 10 x 50 ml culture compared to a 500 ml culture showing that the levels of the expressed lectin decreased during scaling up (see Section 4.1.6). Better aeration obtained in small volumes compared to large volumes could have attributed to the difference in yield. In order to maximize on the yield, 50 ml cultures were pooled for purification by affinity chromatography on mannose-Sepharose-4B.

For analysis of purification, the lectin sample was over-loaded on the SDS-PAGE in order to confirm the purity of the lectin. The distinct bands of lower molecular mass than the expressed lectin are most probably degradation products. This probability is supported by the fact that on Western blot, the lower molecular mass bands were recognized by the antiserum directed against the mukwa seed lectin. Moreover, their intensity increased with the age of the expressed lectin sample. The native lectin also degrades with time but at a much slower rate (S. Beeckmans, personal communication).

The purified lectin has a slightly lower molecular mass than the native lectin from mukwa seeds most probably due to the absence of the C-terminal peptide and the lack of

glycosylation of the recombinant lectin. The native lectin is thought to have a glycosyl moiety since the amino acid residue N¹¹⁸ on the mukwa seed lectin is a potential N-glycosylation site (Loris *et al.*, 2004). Arango and co-workers (1993) showed that the non-glycosylated recombinant *Erythrina corallodendron* lectin had an apparent molecular mass of 28 kDa on SDS-PAGE, about 2 kDa smaller than the glycosylated native lectin. In addition, determination of molecular masses of different native and recombinant legume lectins by mass spectrometry confirmed that glycan structures do contribute to the relative molecular weight (Young *et al.*, 1995).

Since the native mukwa seed lectin and the recombinant one are indistinguishable by specific absorbance, Western blot analysis with anti-mukwa seed lectin polyclonal antibodies and haemagglutinating activity, the recombinant lectin can be used *in vitro* in place of native lectin.

5.2 Mutagenesis of the Sugar-Specificity Loop of the Mukwa Seed Lectin

Having shown that the recombinant wild type mukwa seed lectin was biologically active, we set out to alter the amino acid sequence of loop D of the mukwa seed lectin. The conserved amino acid residues required for sugar binding were identified as D⁸⁶, G¹⁰⁶, N¹³⁸ and F¹³² in the mukwa seed lectin (Loris *et al.*, 2003) and these were retained in the systematically constructed mutants. The *in vitro* site-directed mutagenesis experiments were carried out to try and identify the determinants of the sugar specificity of this lectin, in particular, and of legume lectins in general.

Clone Muk151QII28 was used as the template for the *in vitro* mutagenesis. The incorporation of a restriction endonuclease recognition site in all the mutagenic primers designed made the identification of putative mutants by restriction endonuclease analysis

on agarose gel possible. Although in some cases the actual nucleotide sequence had to be changed in order to provide a restriction endonuclease recognition site, the specific amino acid sequence was maintained.

In the first set of experiments, the nucleotide sequence coding for seven amino acid residues that constituted the part of the mukwa seed lectin specificity loop that interacts with the sugar was removed. This region was replaced with the respective corresponding DNA segment coding for either nine amino acid residues of the UEA II or ten amino acid residues of the ECorL specificity loops. Studies by Jeyaprakash and co-workers (2004) indicated that variation in the length of the sugar-binding loop could be a strategy for generating ligand specificity. The lectin artocarpin has a higher affinity for mannotriose than dimannose, whereas the reverse is true for the heltuba lectin. The sugar-binding loop in artocarpin is four amino acid residues longer than in heltuba. Variation in loop length was also shown to be responsible for the different blood group specificities of the two homologous lectins from winged beans. The basic winged bean lectin, with the longer specificity loop, binds the blood group A and B antigens whereas the acidic winged bean lectin binds the O antigenic determinant (Manoj *et al.*, 2000).

Mutagenic primers Muk-17 and Muk-18 allowed for the introduction of an *EcoRV* recognition site and a *BspEI* recognition site into clone Muk151QII28 at the 3' and 5' ends of the specificity loop respectively. Digestion of the corresponding mutant with these enzymes made it possible for the nucleotide sequence coding for the specificity loop of the mukwa seed lectin to be removed and replaced by the respective UEA II and ECorL specificity loops.

The DNA segment specifying for the UEA II and ECorL specificity loops was obtained by annealing equimolar concentrations of complimentary primers Muk-19 and Muk-20, and Muk-21 and Muk-22 respectively. Introduction of the UEA II specificity

loop into the *EcoRV* recognition site, however, resulted in the phenylalanine residue USP9 being substituted for by tyrosine. Studies by Zhu and co-workers (1996) demonstrated that replacement of the conserved loop C Tyr with Phe in the *Griffonia simplicifolia* lectin II did not affect sugar-binding. The expectation was that the substitution of Phe by Tyr, another aromatic residue in the UEA II mutant would have no effect on sugar binding. On the other hand, introduction of the ECorL specificity loop into the *EcoRV* recognition site gave the amino acid residue ESP10 as aspartic acid (D) instead of the required alanine (A). The nucleotide sequence coding for the D residue was then changed to the nucleotide sequence coding for an A residue using mutagenic primer Muk-35. The presence of the charged D residue instead of the small aliphatic A residue would presumably affect non-polar interaction and most likely abolish sugar-binding (Zhu *et al.*, 1996).

In the second set of experiments, specific mutations, predicted from X-ray crystallography to change the mukwa seed lectin sugar specificity from α -mannose/glucose to β -mannose/glucose, were introduced into clone Muk151QII28. The presence of the specificity loop E²²¹ amino acid residue sterically prevents a β -linkage from being accommodated in the binding site of the mukwa seed lectin (Loris *et al.*, 2004). Consequently, in one mutant, mutagenic primer Muk-30 allowed for the deletion of the nucleotide sequence encoding E²²¹ amino acid residue and the substitution of the nucleotide sequence encoding Q²²² amino acid residue with a G residue in the mukwa seed lectin specificity loop. In the second mutant, mutagenic primer Muk-33 allowed for the substitution of the nucleotide sequence encoding both the E²²¹ and Q²²² residues in the mukwa seed lectin specificity loop with two G residues.

Studies by Loris and co-workers (2003) suggest that although the backbone conformation of both the short and long versions of loop C are not a determinant of monosaccharide specificity, specific side chains on this loop do influence the nature of the

sugar that can be accommodated in the binding site. Consequently, the exact sequence of the metal-binding loop cannot be ignored. For instance, studies by Sharma and co-workers (1998) suggest that the presence of metal-binding loop C Glu¹²⁹ in the peanut agglutinin (PNA) is important in imparting exclusive galactose-specificity in PNA. In another study, the chimeric lectin from the β -galactose-specific *Bauhinia purpurea* agglutinin (BPA) and the α -mannose-specific *Lens culinaris* agglutinin (LCA) showed a distinctly different sugar-binding specificity than that of either BPA or LCA. The chimeric lectin gene had been constructed by replacing the nucleotide sequence coding for nine amino acid residues in the metal-binding region in the cDNA encoding BPA with the corresponding region from LCA (Yamamoto *et al.*, 2000a). Moreover, random mutations introduced in loop C of BPA led to changes in sugar specificity (Yamamoto *et al.*, 2000b).

The chitobiose-specific UEA II can accommodate a β -linkage because of the presence of the bulky Y¹³⁵ amino acid residue in its metal-binding loop (Loris *et al.*, 2000). In both β -mannose/glucose-specific mutants, mutagenic primer Muk-34 therefore allowed for the substitution of the nucleotide sequence encoding S¹³⁷ amino acid residue of the mukwa seed lectin with that of Y. The incorporation of the bulky Y residue was so as to prevent the formation of an α -linkage in the sugar binding site of the mutated protein (Loris *et al.*, 2004).

The *in vitro* mutagenesis experiments therefore resulted in the generation of four clones that were used as the respective template DNA for the UEA II specificity loop, ECorL specificity loop, S¹³⁷ $\rightarrow$ Y + E²²¹Q²²² $\rightarrow$ single G substitution and S¹³⁷ $\rightarrow$ Y + E²²¹Q²²² $\rightarrow$ double G substitution mutants. The DNA region carrying these mutations was then sub-cloned into clone pGV4977, the pBADMyHisA-wild type mukwa seed lectin recombinant in which the corresponding DNA region had been excised. Excision was made possible because restriction endonucleases *Cla*I and *Avr*II have unique recognition

sites in the DNA coding for the mukwa seed lectin (see Appendix C) which are absent in both pUC18 and pBADMycHisA vectors.

Sequencing of the pBADMycHisA-mukwa seed lectin mutants confirmed that only the desired mutations had been incorporated into the coding region of the mukwa seed lectin. The four mutants were expressed in *E. coli* TOP10 cells as had been done for the wild type lectin. Maximum expression of the mutants, as shown by SDS-PAGE, was obtained after induction with 0.2% L-arabinose and after growing the cultures overnight just as for the wild type lectin.

Although the UEA II specificity loop mutant did not cross-react with antiserum directed against the mukwa seed lectin, as evidenced by the agarose gel immunodiffusion tests, SDS-PAGE analysis had shown that this mutant was actually expressed. Western blots to assay if the UEA II specificity loop mutant was capable of binding any glycoproteins were successfully immunoprobed with antiserum directed against the native mukwa seed lectin. Attempts were therefore made to purify this lectin, along with the other recombinant lectins, by affinity chromatography on an anti-native mukwa seed lectin antibody column. The UEA II specificity loop mutant could not be isolated as anticipated suggesting that this mutant was probably being expressed in too low quantities. Higher yields of the mutant lectin might be obtained if specific antibodies against the mutant lectin were developed.

The yield of the wild type mukwa seed recombinant lectin obtained on the IgG-Sepharose-4B column was approximately 25% of that obtained on the mannose-Sepharose-4B column. The lower yield shows that the antibody column was over-loaded. However, no efforts to maximize the isolation of the recombinant lectins were made, as the aim at this stage was to check if pure lectins could be obtained using the antibody column. Lower yields of the mutant lectins may have been obtained on the anti-mukwa seed lectin

IgG-column as the antibodies were not specific for the mutant lectins.

The mutant lectins could have been expressed with a histidine tag at the carboxy-terminus of pBADMycHisA to ease the purification process. For instance, ProBond resin binds hexahistidine-tagged proteins specifically and this in turn increases the yield of purified protein (Invitrogen Life Technologies, 2000).

Analysis of purification of the recombinant lectins on the antibody column showed extra bands of lower molecular mass that were attributed to degradation products. Samples were over-loaded so as to confirm the purity of the recombinant lectins.

Although the mutant lectins were capable of weakly binding to some glycoproteins as evidenced by the Western blots, none of purified mutant lectins, namely the ECorL specificity loop, the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ single G substitution and the $S^{137} \rightarrow Y + E^{221}Q^{222} \rightarrow$ double G substitution mutants agglutinated any of the different animal erythrocytes tested. The complete loss of haemagglutinating activities of these mutants might be due to the conformational change of the sugar-binding region. While oligonucleotide-directed mutagenesis of loop D in the mukwa seed lectin did not make the mutant lectins devoid of carbohydrate-binding activity, the loss of agglutinating activity may have been due to the change in affinity and/or specificity (Chen *et al.*, 2002) or due to failure to form dimers. Enzyme-linked immunosorbent assays (ELISA) using plates that are coated with a range of sugars could possibly have been used to identify the specific sugar(s) that each mutant lectin is capable of binding.

The actual conformation of the lectin mutants is probably not as had been predicted by computer modeling as evidenced by the failure to agglutinate any of the animal erythrocytes tested. For instance, the galactose-specific ECorL agglutinates sheep erythrocytes yet the ECorL specificity loop mutant was not capable of agglutinating these cells. While X-ray diffraction correlates with other physical measurements that are based

on proteins in solution, such as NMR spectroscopy, what is observed in the crystalline state is an average structure of a molecule in which atoms are normally undergoing rapid fluctuations in solution. Consequently, the average crystalline structure may not necessarily be the active structure of a particular protein in solution (Devlin, 1992; Díaz-Maurino *et al.*, 1998). For instance, mutational studies on ECorL showed that although crystal data implicated loop D amino acid residues A²¹⁸ and Q²¹⁹ in sugar-binding, the latter residue was not involved in galactose-binding in solution (Adar and Sharon, 1996).

While slight changes introduced by mutagenesis can change the preference for sugar binding from mannose to galactose (Drickamer, 1995), in this case seven amino acid residues in mukwa seed lectin specificity loop were replaced by ten from the ECorL specificity loop. The number of amino acid residues changed is rather too high to maintain the way the protein folds because the sequence of amino acids ultimately determine the protein's tertiary structure. Non-covalent forces act on the primary structure and cause a protein to fold into a unique conformational structure (Devlin, 1992; Young and Oomen, 1992). Moreover, interactions between the lectin and sugar are limited such that slight modifications of the sugar-binding site can result in dramatic changes in the binding selectivity (Drickamer, 1995). Van Damme and co-workers (1995a) showed that a non-agglutinating lectin-related protein from *Cladrastis lutea* (yellow wood) lacked sugar-binding activity due to the insertion of three extra amino acids in the carbohydrate-binding site. Given that the size of loop D is invariant in mannose/glucose-specific lectins, it is highly likely that the insertion of the ECorL specificity loop amino acid residues into the mukwa seed lectin amino acid framework disrupted the proper folding of the lectin.

5.3 Conclusions

The study showed that it is possible to express the mukwa seed lectin in *E. coli* as a soluble protein using the pBAD expression system. Maximum expression is obtained using 0.2% L-arabinose in cultures grown overnight. Approximately 15 mg of the expressed lectin can be isolated per litre of culture by affinity chromatography on a mannose-Sepharose-4B column. The expressed lectin, about 25 kDa in size, is essentially indistinguishable from the native lectin isolated from mukwa seeds in terms of its biological activities.

The mutagenesis experiments showed that changing the amino acid sequence of loop D of the mukwa seed lectin causes the mutant lectins to lose haemagglutinating activity. It follows that simply exchanging specificity loops cannot produce lectins with desired affinities and specificities. In other words, loop D is not the sole determinant of sugar specificities in legume lectins.

5.4 Recommendations

Analytical isoelectric focusing of the mukwa seed lectin shows at least eight bands upon Coomassie blue staining, with the strongest band having a pI of 6.2 (S. Beeckmans, unpublished). It would be interesting to determine which one of these bands the expressed lectin corresponds to. The stability of the expressed lectin can be studied under different conditions, particularly if the lectin is to be used in histochemistry. The finer specificities of the wild type recombinant lectin and the native lectin can be compared using a range of glycoconjugates.

The UEA II specificity loop mutant that was expressed in low quantities in *E. coli* is most probably toxic to the host cells. It would therefore be worthwhile expressing it in a

eukaryotic expression system such as yeast and determine whether higher yields can be obtained.

The specific sugar(s) that each mutant lectin recognizes can be identified by affinity chromatography. A range of specific sugars can be immobilized and the sugar-Sepharose columns used to try and isolate the mutant lectins and hence identify the specific sugars recognized by the mutant lectins. If a given mutant lectin happens to bind to a particular immobilized sugar, the bound lectin can be eluted using acetic acid. The mutants can be used in comparative analysis of sugar-binding using ELISA and plasmon resonance technology.

X-ray crystallography of the recombinant lectins should be carried out and their three-dimensional structures compared with the computer models. The actual structures of the recombinant lectins can then be super-imposed onto the structures of the native mukwa seed lectin, ECorL and the UEA II to try and identify the amino acid residues that still need to be altered in order to change sugar specificity.

6. REFERENCES

- Adar, R., Moreno, E., Streicher, H., Karlsson, K-A., Angstrom, J. and Sharon, N. (1998). Structural features of the combining site region of *Erythrina corallodendron* lectin: role of tryptophan 135. *Protein Science* **7**, 52-63.
- Adar, R., Richardson, M., Lis, H. and Sharon, N. (1989). The amino acid sequence of *Erythrina corallodendron* lectin and its homology with other legume lectins. *FEBS Letters* **257**, 81-85.
- Adar, R., and Sharon, N. (1996). Mutational studies of the amino acid residues in the combining site of *Erythrina corallodendron* lectin. *European Journal of Biochemistry* **239**, 668-674.
- Amersham Biosciences, UK. (2000). Unique site elimination (U.S.E.) mutagenesis kit. Amersham Biosciences.
- Arango, R., Adar, R., Rozenblatt, S. and Sharon, N. (1992). Expression of *Erythrina corallodendron* lectin in *Escherichia coli*. *European Journal of Biochemistry* **205**, 575-581.
- Arango, R., Rodriguez-Arango, E., Adar, R., Belenky, D., Loontjens, F.G., Rozenblatt, S. and Sharon, N. (1993). Modification by site-directed mutagenesis of the specificity of *Erythrina corallodendron* lectin for galactose derivatives with bulky substituents at C-2. *FEBS Letters* **330**, 133-136.
- Arango, R., Rozenblatt, S. and Sharon, N. (1990). Cloning and sequence analysis of the *Erythrina corallodendron* lectin cDNA. *FEBS Letters* **264** (1), 109-111.
- Ayouba, A., Chatelain, C. and Rouge, P. (1991). Legume lectins interact with muramic acid and N-acetylmuramic acid. *FEBS Letters* **289**, 102-104.
- Banerjee, R., Mande, S.C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S.K., Suguna, K., Surolia, A. and Vijayan, M. (1994). Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *Proceedings of National Academy of Science USA* **91**, 227-231.
- Barbieri, L., Batelli, G.B. and Stirpe, F. (1993). Ribosome-inactivating proteins from plants. *Biochimica et Biophysica Acta* **1154**, 237-282.
- Barondes, S.H. (1988). Bifunctional properties of lectins: lectins redefined. *Trends in Biochemical Sciences* **13**, 480-482.
- Barre, A., Bourne, Y., Van Damme, E.J.M., Peumans, W.J. and Roug  , P. (2001). Mannose-binding plant lectins: different structural scaffolds for a common sugar-recognition process. *Biochimie* **83**, 645-651.

- Baumgartner, P., Raemaekers, R.J.M., Durieux, A., Gatehouse, A., Davies, H. and Taylor, M. (2002). Large-scale production, purification, and characterisation of recombinant *Phaseolus vulgaris* phytohemagglutinin E-form expressed in the methylotrophic yeast *Pichia pastoris*. *Protein Expression and Purification* **26**, 394-405.
- Beeckmans, S. (1999). Chromatography methods to study protein-protein interaction. *Methods* **19**, 278-305.
- Bies, C., Lehr, C.-M. and Woodley, J.F. (2004) Lectin-mediated drug targeting: history and applications. *Advanced Drug Delivery Reviews* **56**, 425-435.
- Bonnadel, J., Sundar, L., Waesche, A. and Greener, A. (2000). Increase transformation efficiencies with optimized electroporation-competent cells. *Stratagene* **13**, 146-147.
- Boorman, K.F., Dodd, B.E. and Lincoln, P.J. (1988). An introduction to blood groups. In: *Blood Group Serology* **6**. Churchill Livingstone, London.
- Bostwick, D.E., Dannehofer, J.M., Skaggs, M.I., Lister, R.M., Larkins, B.A. and Thompson, G.A. (1992). Pumpkin phloem lectin genes are specifically expressed in companion cells. *Plant Cell* **4**, 1539-1548.
- Bouckaert, J., Hamelryck, T., Wyns, L. and Loris, R. (1999). Novel structures of plant lectins and their complexes with carbohydrates. *Current Opinion in Structural Biology* **9**, 572-577.
- Boyd, W.C. and Ruguera, R.M. (1949). Studies on haemagglutinins present in seeds of some representatives of Leguminosae. *Journal of Immunology* **62**, 33-339.
- Boyd, W.C. and Shyleigh, E. (1954). Specific precipitating activity of plant agglutinins (lectins). *Science* **119**, 419.
- Brewin, N.J. and Kardailsky, I.V. (1997). Legume and nodulation by *Rhizobium*. *Trends in Plant Science* **2**, 92-98.
- Brooks, S.A., Leathem, A.J.C. and Schumacher, U. (1997). *Lectin Histochemistry: A Concise Practical Handbook*. Bios Scientific Publishers, Oxford.
- Calvete, J.J., Thole, H.H., Raida, M., Urbanke, C., Romero, A., Grangeiro, T.B., Ramos, M.V., Almeida da Rocha, I.M. Guimaraes, F.N. and Cavada, B.S. (1999). Molecular characterization and crystallization of Diocleinae lectins. *Biochimica et Biophysica Acta* **1430**, 367-375.
- Cammue, B.P.A., Peeters, B. and Peumans, W.J. (1986). A new lectin from tulip (*Tulipa*) bulbs. *Planta* **169**, 583-588.
- Carrington, D.M., Auffret, A. and Hanke, D.E. (1985). Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature* **313**, 64-67.

- Chacko, B.K. and Appukuttan, P.S. (2001). Peanut (*Arachis hypogaea*) lectin recognizes α -linked galactose, but not N-acetyllactosamine in N-linked oligosaccharide terminals. *International Journal of Biological Macromolecules* **28**, 365-371.
- Chapot, M.P., Peumans, W.J. and Strosberg, A.D. (1986). Extensive homologies between lectins from non-leguminous plants. *FEBS Letters* **195** (1,2), 231-234.
- Chen, Y., Rouge, P., Peumans, W.J. and Van Damme, E.J.M. (2002). Mutational analysis of the carbohydrate-binding activity of the NeuAc(α -2,6)Gal/GalNAc-specific type 2 ribosome-inactivating protein from elderberry (sambucus nigra) fruits. *Biochemistry Journal* **364**, 587-592.
- Chrispeels, M.J. and Raikhel, N.V. (1991). Lectins, lectin genes, and their role in plant defense. *The Plant Cell* **3**, 1-9.
- Colson, C., Glover, S.N., Symonds, N. and Stacey, K.A. (1965). The location of the genes for host-controlled modification and restriction in *Escherichia coli* K-12. *Genetics* **52**, 1043-1050.
- Conley, E.C. and Saunders, J.R. (1984). Recombination-dependent recircularization of linearized pBR322 plasmid DNA following transformation of *Escherichia coli*. *Molecular and General Genetics* **194**, 211-218.
- Devlin, T.M. (1992). *Textbook of Biochemistry with Clinical Correlations*. 3rd Edition. John Wiley and Sons, New York, pp. 25-88.
- Díaz, C., Melchers, L.S., Hooykaas, P.J.J., Lutenberg, B.J.J. and Kijne, J.W. (1989). Root lectin as a determinant of host plant specificity in the *Rhizobium*-legume symbiosis. *Nature* **338**, 579-581.
- Díaz-Maurino, T., Solis, D., Jimenez-Barbero, J., Martin-Lomas, M., Siebert, H. and Vliegthart, J.F.G. (1998). Lectin meets ligand. *Carbohydrates in Europe* **23**, 36-41.
- Drickamer, K. (1988). Two distinct classes of carbohydrate-recognition domains in animal lectins. *Journal of Biological Chemistry* **263**, 9557-9560.
- Drickamer, K. (1995). Multiplicity of lectin-carbohydrate interactions. *Nature Structural Biology* **2**(6), 437-439.
- Drickamer, K. (1997). Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins. *Structure* **5**, 465-468.
- Duranti, M. and Gius, C. (1997). Legume seed: protein content and nutritional value. *Field Crops Research* **53**, 31-45.
- Echemendia-Blanco, D. (2001). Purification and characterization of lectins from *Pterocarpus* species. MSc. Thesis, Interuniversity Program Molecular Biology, Belgium.

- Edelman, G.M., Cunningham, B.A., Reeke, G.N., Becker, J.W., Waxdal, M.J. and Wang, J.L. (1972). The covalent and three-dimensional structure of concanavalin A. *Proceedings of the National Academy of Science USA* **69**, 2580-2584.
- Elgavish, S. and Shaanan, B. (1997). Lectin-carbohydrate interactions: different folds, common recognition principles. *Trends in Biochemical Sciences* **22**, 462-467.
- Etzler, M.E. (1985). Plant lectins: molecular and biological aspects. *Annual Reviews in Plant Physiology* **36**, 209-234.
- Etzler, M.E. (1986). Distribution and function of plant lectins. In: *The Lectins: Properties, Function and Applications in Biology and Medicine*. Liener, I.E., Sharon, N. and Goldstein, I.J. (Eds). Academic Press, Orlando, pp. 371-435.
- Etzler, M.E. (1992). Plant lectins: molecular biology, synthesis and function. In: *Glycoconjugates: Composition, Structure and Function*. Allen, H.J. and Kisailus, E.C. (Eds). Marcel Dekker, New York, pp. 521-539.
- Ewen, S.W.B. (1998). Lectin histochemistry in histology and histopathology. In: *Handbook of Plant Lectins: Properties and Biochemical Applications*. Van Damme, E.J.M., Peumans, W.J., Pusztai, A. and Bardocz, S. (Eds). John Wiley and Sons, New York, pp. 51-59.
- Ewen, S.W.B. and Pusztai, A. (1999). Effect of diets containing genetically modified potatoes expressing *Galanthus nivalis* lectin on rat small intestine. *The Lancet* **354**, 353-1354.
- Foissac, X., Thi Loc, N., Christou, P., Gatehouse, A.M.R. and Gatehouse, J.A. (2000). Resistance to green leafhopper (*Nephotettix virescens*) and brown planthopper (*Nilaparvata lugens*) in transgenic rice expressing snowdrop lectin (*Galanthus nivalis* agglutinin; GNA). *Journal of Insect Physiology* **46**, 573-583.
- Foriers, A., Lebrun, A., Van Rapenbusch, R., De Neve, R. and Strosberg, A.D. (1981). The structure of the lentil (*Lens culinaris*) lectin. Amino acid sequence determination and prediction of secondary structure. *Journal of Biological Chemistry* **256**, 5550-5560.
- Franz, H. (1988). The ricin story. In: *Advances in Lectin Research* Volume 1. Franz, H. (Ed). Verlag Volk and Gesundheit, Berlin, pp. 10-25.
- Freire, M.G., Gomes, V.M., Corsini, R.E., Machado, O.L.T., De Simone, S.G., Novello, J.C., Marangoni, S. and Macedo, M.L.R. (2002). Isolation and partial characterization of a novel lectin from *Talisia esculenta* seeds that interferes with fungal growth. *Plant Physiology and Biochemistry* **40**, 61-68.
- Funston, M. (1993). *Bushveld trees: Lifeblood of the Transvaal lowveld*. Fernwood Press, Cape Town.

- Gabius, H.J. and Gabius, S. (1991) Tumour lectinology: a glycobiological approach to tumour diagnosis and therapy. In: *Lectin Review Volume 1*. Kilpatrick, D.C., Van Driessche, E. and Bøgg-Hansen, T.C. (Eds). Sigma Chemical Co., USA, pp. 91-102.
- Gatehouse, A.M.R., Barbieri, L., Stirpe, F. and Croy, R.R.D. (1990). Effects of ribosome inactivating proteins on insect development – differences between Lepidoptera and Coleopteran. *Entomologia Experimentalis et Applicata* **54**, 43-51.
- Gohier, A., Espinosa, J. F., Jimenez-Barbero, J., Carrupt, P.-A., Perez, S. and Imberty, A. (1996). Knowledge-based modeling of a legume lectin and docking of the carbohydrate ligand: The *Ulex europaeus* lectin I and its interaction with fucose. *Journal of Molecular Graphics* **14**, 322-327.
- Goldstein, I.J. and Hayes, C.E. (1978). The lectins: carbohydrate-binding proteins of plants and animals. *Advances in Carbohydrate Chemistry and Biochemistry* **35**, 127-340.
- Goldstein, I.J., Hughes, R.C., Monsigny, M., Osawa, T. and Sharon, N. (1980). What should be called a lectin? *Nature* **285**, 66.
- Goldstein, I.J. and Poretz, R.D. (1986). Isolation, physicochemical characterization, and carbohydrate-binding specificity of lectins. In: *The Lectins: Properties, Function and Applications in Biology and Medicine*. Liener, I.E., Sharon, N. and Goldstein, I.J. (Eds). Academic Press, Orlando, pp. 35-247.
- Gottesman, S. (1996). Proteases and their targets in *Escherichia coli*. *Annual Reviews of Genetics* **30**, 465-506.
- Grant, G., More, L.J., Mckenzie, N.H., Dorward, P.M., Buchan, W.C., Telek, L. and Pusztai, A. (1995). Nutritional and haemagglutination properties of several tropical seeds. *Journal of Agricultural Science* **124**, 437-445.
- Green, A.A. and Hughes, C.E. (1955). Protein fractionation on the basis of solubility in aqueous solution of salts and organic solvents. In: *Methods in Enzymology* Volume **1**. Colowick, S.P. and Kaplan, N.O. (Eds), pp. 67-90.
- Guzman, L-M., Belin, D., Carson, M.J. and Beckwith, J. (1995). Tight regulation, modulation and high-level expression by vectors containing the arabinose P_{BAD} promoter. *Journal of Bacteriology* **177 (14)**, 4121-4130.
- Guzmán-Partida, A.M., Robles-Burgueno, M.R., Ortega-Nieblas, M. and Vázquez-Moreno, I. (2004). Purification and characterization of complex carbohydrate specific isolectins from wild legume seeds: *Acacia constricta* (vinorama) is highly homologous to *Phaseolus vulgaris* lectins. *Biochimie* **86**, 335-342.
- Hamelryck, T.W., Loris, R., Bouckaert, J. and Wyns, L. (1998). Structural features of the legume lectins. *Trends in Glycoscience and Glycotechnology* **10 (55)**, 349-360.

- Hamelryck, T.W., Loris, R., Bouckaert, J., Dao-Thi, M-H., Strecker, G., Imberty, A. Fernandez, E., Wyns, L. and Etzler, M.E. (1999). Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. *Journal of Molecular Biology* **286**, 1161-1177.
- Hanahan, D., Jessee, J. and Bloom, F.R. (1991). Plasmid transformation of *Escherichia coli* and other bacteria. In: *Methods in Enzymology* Volume **204**, Bacterial Genetic Systems. Miller, J.H. (Ed). Academic Press, New York, pp. 63-113.
- Hardman, K.D. and Ainsworth, C.F. (1972). Structure of concanavalin A at 2.4 Å resolution. *Biochemistry* **11**, 4910-4919.
- Higgins, T.J.V., Chandler, P.M., Zurawski, G., Button, S.C. and Spencer, D. (1983a). The biosynthesis and primary structure of pea seed lectin. *The Journal of Biological Chemistry* **258** (15), 9544-9549.
- Higgins, T.J.V., Chrispeels, M.J., Chandler, P.M. and Spencer, D. (1983b). Intracellular sites of synthesis and processing of lectin in developing pea cotyledons. *The Journal of Biological Chemistry* **258** (15), 9550-9552.
- Hilder, V.A., Powell, K.S., Gatehouse, J.A., Gatehouse, L.M., Shi, Y., Hamilton, W.D.O., Merryweather, A., Newell, C.A., Timans, J.C., Peumans, W.J., Van Damme, E.J.M. and Boulter, D. (1995). Expression of snowdrop lectin in transgenic tobacco plants results in added protection against aphids. *Transgenic Research* **4**, 18-25.
- Hitti, Y.S. and Bertino, A.M. (1994). Proteinase K and T4 DNA polymerase facilitate the blunt-end subcloning of PCR products. *BioTechniques* **16**, 803-805.
- Hoedemaeker, F.J., Richardson, M., Díaz, C.L., De Peter, S.B. and Kijne, J.W. (1994). Pea (*Pisum sativum* L.) seed isolectins 1 and 2 and pea root lectin result from carboxypeptidase-like processing of a single gene product. *Plant Molecular Biology* **24**, 75-81.
- Hoffman, L.M. and Donaldson, D.D. (1985). Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. *The EMBO Journal* **4**, 883-889.
- Iglesias, J.L., Lis, H. and Sharon, N. (1982). Purification and properties of a D-galactose/N-acetyl-D-galactosamine-specific lectin from *Erythrina cristagalli*. *European Journal of Biochemistry* **123**, 247-252.
- Invitrogen Life Technologies, USA. (2000). pBAD/His A, B and C, pBAD/MycHis A, B and C, Version C. Invitrogen Corporation.
- Jeyaprakash, A.A., Srivastav, A., Surolia, A. and Vijayan, M. (2004). Structural Basis for the carbohydrate specificities of artocarpin: variation in the length of a loop as a strategy for generating ligand specificity. *Journal of Molecular Biology* **338**, 757-770.

- Jordan, E.T., and Goldstein, I.J. (1994). The sequence of a second member of the lima bean lectin gene family and the expression and characterization of recombinant lectin in *Escherichia coli*. *The Journal of Biological Chemistry* **269** (10), 7674-7681.
- Kalsi, G., Babu, C.R. and Das, R.H. (1995). Localization of peanut (*Arachis hypogaea*) root lectin (PRA II) on root surface and its biological significance. *Glycoconjugate Journal* **12**, 45-50.
- Kaul, R., Read, J. and Mattiasson, B. (1991). Screening for plant lectins by latex agglutination tests. *Phytochemistry* **30**(12), 4005-4009.
- Kennedy, J.F., Palva, P.M.G., Corella, M.T.S., Cavalcanti, M.S.M. and Coelho, L.C.B.B. (1995). Lectins, versatile proteins of recognition: a review. *Carbohydrate Polymers* **26**, 219-230.
- Kenoth, R., Komath, S.S. and Swamy, M.J. (2003). Physicochemical and saccharide-binding studies on the galactose-specific seed lectin from *Trichosanthes cucumerina*. *Archives of Biochemistry and Biophysics* **413**, 131-138.
- Kilpatrick, D.C. (1995). Lectins in immunology. In: *Lectins: Biomedical Perspectives*. Pusztai, A. and Bardocz, S. (Eds). Taylor and Francis, London, pp. 155-178.
- Kilpatrick, D.C. (2002). Animal lectins: a historical introduction and overview. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1572**, 187-197.
- Kobata, A. and Endo, T. (1992). Immobilized lectin columns: useful tools for the fractionation and structural analysis of oligosaccharide. *Journal of Chromatography* **597**, 111-122.
- Kocourek, J. and Horejši, V. (1983). A note on the recent discussion on definition of the term 'lectin'. In: *Lectins: Biology Biochemistry and Clinical Biochemistry* Volume 3. Bøg-Hansen, T.C. and Spengler, G.A. (Eds). Walter de Gruyter, Berlin, pp. 3-6.
- Kohn, J. and Wilchek, M. (1984). The use of cyanogen bromide and other novel cyanylating agents for the activation of polysaccharide resins. *Applied Biochemistry and Biotechnology* **9**, 285-305.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of proteins from the head of bacteriophage T4. *Nature* **227**, 680-685.
- Laursen, B.S., Sørensen, H.P., Mortensen, K.K. and Sperling-Petersen, H.U. (2005). Initiation of protein synthesis in bacteria. *Microbiology and Molecular Biology Reviews* **69**(1), 101-123.
- Law, I.J. (2000). Comparison of putative nodule and seed lectin gene promoters of peanut and *in situ* localization of nodule lectin gene expression. *Plant Science* **153**, 43-54.

- Liener, I.E., (1986). Nutritional significance of lectins in the diet. In: *The Lectins: Properties, Function and Application in Biology and Medicine*. Liener, I.E., Sharon, N. and Goldstein, I.R. (Eds). Academic Press, London, pp. 527-552.
- Lilie, H., Schwarz, E. and Rudolph, R. (1998). Advances in refolding of proteins produced in *E. coli*. *Current Opinion in Biotechnology* **9**, 497-501.
- Lis, H. and Sharon, N. (1986). Lectins as molecules and as tools. *Annual Reviews in Biochemistry* **55**, 35-67.
- Lis, H. and Sharon, N. (1998). Lectins: carbohydrate specific proteins that mediate cellular recognition. *Chemistry Reviews* **98**, 637-674.
- Lord, J.M. (1985). The structure and synthesis of plant lectins. *The New Phytologist* **101**, 351-366.
- Loris, R. (2002). Principles of structures of animal and plant lectins. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1572**, 198-208.
- Loris, R., De Greve, H., Dao-Thi, M-H., Messens, J., Imberty, A. and Wyns, L. (2000). Structural basis of carbohydrate recognition by lectin II from *Ulex europaeus*, a protein with a promiscuous carbohydrate-binding site. *Journal of Molecular Biology* **301**, 987-1002.
- Loris, R., Hamelryck, T., Bouckaert, J. and Wyns, L. (1998). Legume lectin structure. *Biochimica et Biophysica Acta* **1383**, 9-36.
- Loris, R., Imberty, A., Beeckmans, S., Van Driessche, E., Read, J.S., Bouckaert, J., De Greve, H., Buts, L. and Wyns, L. (2003). Crystal structure of *Pterocarpus angolensis* lectin in complex with glucose, sucrose and turanose. *The Journal of Biological Chemistry* **278** (18), 16297-16303.
- Loris, R., Van Walle, I., De Greve, H., Beeckmans, S., Deboeck, F., Wyns, L. and Bouckaert, J. (2004). Structural basis of oligomannose recognition by the *Pterocarpus angolensis* seed lectin. *Journal of Molecular Biology* **335**, 1227-1240.
- Machuka, J.S., Okeola, O.G., Van Damme E.J.M., Chrispeels, M.J., Van Leuven, F. and Peumans, W.J. (1999). Isolation and partial characterisation of galactose-specific lectins from African yam beans, *Sphenostyles stenocarpa* Harms. *Phytochemistry* **51**, 721-728.
- Manoj, N., Srinivas, V.R., Surolia, A., Vijayan, M. and Suguna, K. (2000). Carbohydrate specificity and salt-bridge mediated conformational change in acidic winged bean agglutinin. *Journal of Molecular Biology* **302**, 1129-1137.
- Manyumwa, L. K. (1998). Application of a lectin purified from mukwa (*Pterocarpus angolensis*) seed in protein chemistry, cytochemistry and histochemistry. MSc Thesis, University of Zimbabwe.

- Maramba, E. (1998). Purification and characterisation of *Pterocarpus angolensis* (mukwa) lectin. MSc Thesis, University of Zimbabwe.
- Messing, J. (1979). A multipurpose cloning based on single-stranded DNA bacteriophage M13. *Recombinant DNA Technical Bulletin* **2** (2), 43.
- Mirkov, T.E., Wahlstrom, J.M., Hagiwara, K., Finardi-Filho, F., Kjemtrup, S. and Chrispeels, M.J. (1994). Evolutionary relationship among proteins in the phytohemagglutinin arcelin- α -amylase inhibitor family of the common bean and its relatives. *Plant Molecular Biology* **26**, 1103-1113.
- Moore, H.H. (1979). A study of trees and shrubs indigenous to Rhodesia for the presence of human blood group specific haemagglutinins. MPhil Thesis, University of Rhodesia.
- Moore, J.G., Fuchs, C.A., Hata, Y.S., Hicklin, D.J., Colucci, G., Chrispeels, M.J. and Feldman, M. (2000). A new lectin in red kidney beans called PvFRIL stimulates proliferation of NIH 3T3 cells expressing the Flt3 receptor. *Biochimica et Biophysica Acta* **1475**, 216-224.
- Moreno, E., Teneberg, S., Adar, R., Sharon, N., Karlsson, K-A. and Angstrom, J. (1997). Redefinition of the carbohydrate specificity of *Erythrina corallodendron* lectin based on solid-phase binding assays and molecular modelling of native and recombinant forms obtained by site-directed mutagenesis. *Biochemistry* **36**, 4429-4437.
- Neumann, D., Lehr, C.-M., Lenhof, H.-P. and Kohlbacher, O.U. (2004). Computational modeling of the sugar-lectin interaction. *Advanced Drug Delivery Reviews* **56**, 437-457.
- Nishiguchi, M., Yoshida, K., Sumizono, T. and Tazaki, K. (1997). Studies by site-directed mutagenesis of the carbohydrate-binding properties of a bark lectin from *Robinia pseudoacacia*. *FEBS Letters* **403**, 294-298.
- Nomura, K., Ashida, H., Uemura, N., Kushipe, S., Ozaki, T. and Yoshida, M. (1998). Purification and characterization of a mannose/glucose-specific lectin from *Castanea crenata*. *Phytochemistry* **49** (3), 667-673.
- Nollet, H., Deprez, P., Van Driessche, E. and Muylle, E. (1999). Protection of just weaned pigs against infection with F18⁺ *Escherichia coli* by non-immune plasma powder. *Veterinary Microbiology* **65**, 37-45.
- Nowell, P.C. (1960). Phytohaemagglutinin: an initiator of mitosis in cultures of normal human leucocytes. *Cancer Research* **20**, 462-466.
- Palgrave, K.C. (1981). *Trees of Southern Africa*. Struik Publishers, Cape Town.
- Peumans, W.J., Hause, B. and Van Damme, E.J.M. (2000). The galactose-binding and mannose-binding jacalin-related lectins are located in different sub-cellular compartments. *FEBS Letters* **477**, 186-192.

- Peumans, W.J. and Van Damme, E.J.M. (1993). The role of lectins in the defense of plants. In: *Lectins: Biology, Biochemistry, Clinical Biochemistry* Volume **8**. Van Driessche, E., Franz, H., Beeckmans, S., Pfüller, U., Kallikom, A. and Bøg-Hansen, T.C. (Eds). Textop, Hellerup, Denmark, pp. 82-91.
- Peumans, W.J. and Van Damme, E.J.M. (1995). Lectins as plant defense proteins. *Plant Physiology* **109**, 347-352.
- Porath, J. (1974). General methods and coupling procedures. *Methods in Enzymology* Volume **34**, 13-30.
- Pryme, I.F., Bardocz, S., Pusztai, A., Ewen, S.W. and Pfuller, U. (2004). A mistletoe lectin (ML-1)-containing diet reduces the viability of a murine non-Hodgkin lymphoma tumor. *Cancer Detection and Prevention* **28**, 52-56.
- Pusztai, A. (1991). Plant lectins. In: *Chemistry and Pharmacology of Natural Products*. Cambridge University Press, England.
- Pusztai, A. and Bardocz, S. (1997). Biological effects of plant lectins on the gastrointestinal tract: metabolic consequences and application. *Trends in Glycoscience and Glycotechnology* **8**, 149-165.
- Pusztai, A., Grant, G. and Bardocz, S. (1993a). Lectins as metabolic and growth signals: an overview. In: *Lectins: Biology, Biochemistry, Clinical Biochemistry* Volume **8**. Van Driessche, E., Franz, H., Beeckmans, S., Pfüller, U., Kallikom, A. and Bøg-Hansen, T.C. (Eds). Textop, Hellerup, Denmark, pp. 245-249.
- Pusztai, A., Grant, G., Spencer, R.J., Duguid, T.J., Brown, D.S., Ewen, S.W.B., Peumans, W.J., Van Damme, E.J.M. and Bardocz, S. (1993b). Kidney bean lectin-induced *Escherichia coli* overgrowth in the small intestine is blocked by GNA, a mannose-specific lectin. *Journal of Applied Bacteriology* **75**, 360-368.
- Qiagen, Germany. (2000a). Qiaprep miniprep handbook. Qiagen.
- Qiagen, Germany. (2000b). Qiaprep PCR purification handbook. Qiagen.
- Raikhel, N.V., Lee H.-I. and Broekaert, W.F. (1993). Structure and function of chitin-binding proteins. *Annual Reviews in Plant Physiology and Plant Molecular Biology* **44**, 591-615.
- Rao, V.S.R., Lam, K. and Qasba, P.K. (1998). Architecture of the sugar binding sites in carbohydrate binding proteins-a computer modeling study. *International Journal of Biological Macromolecules* **23**, 295-307.
- Rego, E.J.L., de Carvalho, D.D., Marangoni, S., de Oliveira, B. and Novello, J.C. (2002). Lectins from seeds of *Crotalaria pallida* (smooth rattlebox). *Phytochemistry* **60**, 441-446.
- Rini, J.M. (1995). Lectin structure. *Annual Review of Biophysics and Biomolecular Structure* **24**, 551-577.

- Roelcke, D. (1995). Serology, biochemistry and pathology of antigens defined by cold agglutinins. In: *Blood Cell Biochemistry* Volume **6**. Cartron, J-P and Rouger, P. (Eds). Plenum Press, New York, pp. 117-152.
- Rojo, J. (1977). Pantropic speciation of *Pterocarpus* (Leguminosae-Papilionaceae) and the Malasia-pacific species. *Forestry Abstracts* **3(1)**, 1932.
- Rougè, P., Cambillau, C. and Bourne, Y. (1991). The three dimensional structure of legume lectins. In: *Lectin Reviews* Volume **1**. Kiltatric, D.C., Van Driessche, E. and Bøg-Hansen, T.C. (Eds). Sigma Chemical Company, St Louis, pp. 143-159.
- Rougè, P., Peumans, W. J., Barre, A. and Van Damme, E. J. M. (2003). A structural basis for the difference in specificity between the two jacalin-related lectins from mulberry (*Morus nigra*) bark. *Biochemical and Biophysical Research Communications* **304**, 91-97.
- Rüdiger, H. (1988). Preparation of plant lectins. In: *Advances in Lectin Research* Volume **1**. Franz, H. (Ed). Springer, Berlin, pp. 26-72.
- Rüdiger, H. (1997). Structure and function of plant lectins. In: *Glycosciences: Status and Perspectives*. Gabius, H.J. and Gabius, S. (Eds). Chapman and Hall, London, pp.415-438.
- Rüdiger, H. and Rougè, P. (1998). Structure and function of plant lectins. *Carbohydrate in Europe* **23**, 18-22.
- Sambrook, J. and Russell, D.W. (2001). *Molecular Cloning: A Laboratory Manual*. 3rd Edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science USA* **74**, 5463-5467.
- Santos, A.C.O., Peixoto, C.A. and Coelho, L.C.B.B. (2004). Ultrastructural analysis and immunocytochemical localization of isolectins in *Cratylia mollis* seeds. *Micron* **35**, 613-618.
- Schägger, H. and Von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry* **166**, 368-379.
- Schwartz, M.W., Caro, T.M. and Banda-Sakala, T. (2002). Assessing the sustainability of harvest of *Pterocarpus angolensis* in Rukwa Region, Tanzania. *Forest Ecology and Management* **170**, 259-269.
- Seshagirirao, K. and Prasad, M.N.V. (1995). Purification and partial characterization of a lectin from *Euphorbia neriifolia* latex. *Biochemistry and Molecular Biology International* **35(6)**, 1199-1204.

- Shaanan, B., Lis, H. and Sharon, N. (1991). Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science* **254**, 262-265.
- Shackleton, C.M. (2002). Growth patterns of *Pterocarpus angolensis* in savannas of the South African lowveld. *Forest Ecology and Management* **166**, 85-97.
- Sharma, V., Srinivas, V.R., Adhikari, P., Vijayan, M. and Surolia, A. (1998). Molecular basis of recognition by Gal/GalNAc specific legume lectins: influence of Glu129 on the specificity of peanut agglutinin (PNA) towards C2-substituents of galactose. *Glycobiology* **8(10)**, 1007-1012.
- Sharma, V., Srinivas, V.R. and Surolia, A. (1996). Cloning and sequencing of winged bean (*Psophocarpus tetragonolobus*) basic agglutinin (WBA I): presence of second glycosylation site and its implications in quaternary structure. *FEBS Letters* **389**, 289-292.
- Sharma, V. and Surolia, A. (1997). Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. *Journal of Molecular Biology* **267**, 433-445.
- Sharon, N. (1977). Lectins. *Scientific American* **236**, 108-119.
- Sharon, N. (1994). Tools for glycobiology. Oxford Glycosystems, pp. 34-36.
- Sharon, N. and Lis, H. (1989). Lectins as cell recognition molecules. *Science* **246**, 227-234.
- Sharon, N. and Lis, H. (1990). Legume lectins- a large family of homologous proteins. *FASEB Journal* **4**, 3198 - 3207.
- Sharon, N. and Lis, H. (1998). 110 years of lectin research. *Carbohydrates in Europe* **23**, 12-17.
- Sharon, N. and Lis, H. (2003). *Lectins*. 2nd Edition. Kluwer Academic Publishers, Dordrecht, Boston, London, pp. 261-311.
- Shibuya, N., Goldstein, I.J., Broekaert, W.F., Nsimba-Lubaki, M., Peeters, B. and Peumans, W.J. (1987). The elderberry (*Sambucus nigra*) bark lectin recognizes the Neu5Ac (α 2-6)Gal/GalNAc sequence. *Journal of Biological Chemistry* **262**, 1596-1601.
- Singh, J., Singh, J. and Kamboj, S.S. (2004). A novel mitogenic and antiproliferative lectin from a wild cobra lily, *Arisaema flavum*. *Biochemical and Biophysical Research Communications* **318**, 1057-1065.
- Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Garther, F.H., Provenzano, M.D., Figimoto, E.K., Goeke, N.M., Olson, B.J. and Klenk, D.C. (1985). Measurements of proteins using bicinchoninic acid. *Analytical Biochemistry* **150**, 76-85.

- Srinivas, V.R., Acharya, S., Rawat, S., Sharma, V. and Surolia, A. (2000). The primary structure of the acidic lectin from winged bean (*Psophocarpus tetragonolobus*): insights in carbohydrate recognition, adenine binding and quaternary association. *FEBS Letters* **474**, 76-82.
- 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. *Protein Expression and Purification* **30**, 283-292.
- Steinbuch, M. and Audran, R. (1969). The isolation of IgG from mammalian sera with the aid of caprylic acid. *Archives of Biochemistry and Biophysics* **134**, 279-284.
- Stratagene, USA. (2000). BL21-CodonPlus-RIL competent cells. Stratagene.
- Strosberg, A.D., Buffard, D., Lauwereys, M. and Foriers, A. (1986). Legume lectins: a large family of homologous proteins. In: *The Lectins: Properties, Functions and Applications in Biology and Medicine*. Liener, I.E., Sharon, N. and Goldstein, I.R. (Eds). Academic Press, Orlando, pp. 249-264.
- Stubbs, M.E., Carver, J.P. and Dunn, R.J. (1986). Production of pea lectin in *Escherichia coli*. *The Journal of Biological Chemistry* **261** (14), 6141-6144.
- Sumar, N. (1994). Lectins as indicators of disease. Oxford Glycosystems, pp. 38.
- Sumner, J.B. and Howell, S.F. (1936). The identification of the haemagglutinin of the jack bean with concanavalin A. *Journal of Bacteriology* **32**, 227-237.
- Swain, L.J., Ebrahimi, S.A. and Rogers, D.J. (1996). Quantitative determination of haemagglutination activity of lectins. In: *Lectins: Biology, Biochemistry and Clinical Biochemistry* Volume **11**. Van Driessche, E., Rougé, P., Beeckmans, S. and Bøg-Hansen, T.C. (Eds). Textop, Hellerup, Denmark, pp. 112-118.
- Takara Biomedicals, Japan. (2000). TaKaRa Ex TaqTM DNA polymerase. Takara Biomedicals.
- Thomas, C.J. and Surolia, A. (2000). Mode of molecular recognition of L-fucose by fucose-binding legume lectins. *Biochemical and Biophysical Research Communications* **268**, 262-267.
- Towbin, H., Staehelin, T. and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Science USA* **76**, 4350-4354.
- United States Biochemical Corporation, USA. (2000). Thermo-sequenase radiolabelled terminator cycle sequencing kit. United States Biochemical Corporation.

- Van Damme, E.J.M., Barre, A. Berner, V., Roug , P., Van Leuven, F. and Peumans, W.J. (1995a). A lectin and lectin-related protein are the two most prominent proteins in the bark of yellow wood (*Cladrastis lutea*). *Plant Molecular Biology* **29**, 579-598.
- Van Damme, E.J.M., Peumans, W.J., Barre, A. and Roug , P. (1998a). Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Critical Reviews in Plant Sciences* **17(8)**, 575-692.
- Van Damme, E.J.M., Peumans, W.J., Pusztai, A. and Bardocz, S. (1998b). *Handbook of Plant Lectins: Properties and Biomedical Applications*. John Wiley and Sons, Chichester, pp. 1-61.
- Van Damme, E.J.M., Smeets, K. and Peumans, W.J. (1995b). The mannose-binding monocot lectins and their genes. In: *Lectins: Biomedical Perspectives*. Pusztai, A. and Bardocz, S. (Eds). Taylor and Francis, London, pp. 59-80.
- Van Damme, E.J.M., Smeets, K., Torrekens, S., Van Leuven, F. and Peumans, W.J. (1994). Characterization and molecular cloning of mannose-binding lectins from the orchidaceae species *Listera ovata*, *Epipactis helleborine* and *Cymbidium hybrid*. *European Journal of Biochemistry* **221**, 769-777.
- Van Damme, E.J.M., Smeets, K., Van Driessche, E. and Peumans, W.J. (1993). Plant lectins: reflections on recent developments and future prospects. In: *Lectins: Biology, Biochemistry and Clinical Biochemistry* Volume **8**. Van Driessche, E., Franz, H., Beeckmans, S., P  ller, U., Kallikom, A. and B  g-Hansen, T.C. (Eds). Textop, Hellerup, Denmark, pp. 73-81.
- Van Driessche, E. (1988). Structure and function of leguminosae lectins. In: *Advances in Lectin Research* Volume **1**. Franz, H. (Ed). VEB Verlag Volk and Gesundheit, Berlin, pp. 73-134.
- Van Driessche, E., Beeckmans, S., Dejaegere, R. and Kanarek, L. (1988). Isolation of the pea lectin precursor and characterization of its processing products. In: *Lectins: Biology, Biochemistry and Clinical Biochemistry* Volume **6**. B  g-Hansen, T.C. and Freed, D.L.J. (Eds). Sigma Press, St. Louis, pp. 355-362.
- Van Driessche, E., de Cupere, F. and Beeckmans, S. (2000). Protein-carbohydrate interaction in the attachment of enterotoxigenic *Escherichia coli* to the intestinal mucosa. In: *Lectins and Pathology*. Caron, M. and Steve, A-P. (Eds). Harwood Academic Publishers, Australia.
- Van Eijsden, R.R., Hoedemaeker, F.J., Diaz, C.L., Lutenberg, B.J.J., De Pater, B.S. and Kijne, J.W. (1992). Mutational analysis of pea lectin. Substitution of Asn125 for Asp in the monosaccharide-binding site eliminates mannose/glucose binding activity. *Plant Molecular Biology* **20**, 1049-1058.
- Van Wyk, B. and Van Wyk, P. (1997). *Field guide to trees of Southern Africa*. Struik Publishers Ltd., Cape Town.

- Van Wyk, P. (1993). Southern African trees: a photographic guide. 1st Edition. Struik Publishers, Cape Town, pp. 57.
- Vasconcelos, I.M. and Oliveira, J.T.A. (2004). Antinutritional properties of plant lectins. *Toxicon* **44**, 385-403.
- Verma, R., Boleti, E. and George, A.J.T. (1998). Antibody engineering: comparison of bacterial, yeast, insect and mammalian expression systems. *Journal of Immunological Methods* **216**, 165-181.
- Vijayan, M. and Chandra, N. (1999). Lectins. *Current Opinion in Structural Biology* **9**, 707-714.
- Weis, W.I. and Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annual Reviews in Biochemistry* **65**, 441-473.
- Woodley, J.F. and Naisbett, B. (1988). The potential of lectins for delaying the intestinal transit of drugs. *Proceedings of the International Symposium on Controlled Release of Bioactive Material* **15**, 125-126.
- Yamamoto, K., Konami, Y. and Osawa, T. (2000a). A chimeric lectin formed from *Bauhinia purpurea* lectin and *Lens culinaris* lectin recognizes a unique carbohydrate structure. *Journal of Biochemistry* **127**, 129-135.
- Yamamoto, K., Maruyama, I.N. and Osawa, T. (2000b). Cyborg lectins: novel leguminous lectins with unique specificities. *Journal of Biochemistry* **127**, 137-142.
- Yamauchi, D. and Minamikawa, T. (1990). Structure of the gene encoding concanavalin A from *Canavalia gladiata* and its expression in *Escherichia coli* cells. *FEBS Letters* **260(1)**, 127-130.
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.
- Yim, M., Ono, T. and Irimura, T. (2001). Mutated plant lectin library useful to identify different cells. *Proceedings of the National Academy of Science USA* **98 (5)**, 2222-2225.
- Young, N.M. and Oomen, R.P. (1992). Analysis of sequence variation among legume lectins: a ring of hypervariable residues forms the perimeter of the carbohydrate-binding site. *Journal of Molecular Biology* **228**, 924-934.
- Young, N.M., Watson, D.C., Yaguchi, M., Adar, R., Arango, R., Rodriguez-Arango, E., Sharon, N., Blay, P.K.S. and Thibault, P. (1995). C-terminal post-translational proteolysis of plant lectins and their recombinant forms expressed in *Escherichia coli*. *The Journal of Biological Chemistry* **270 (6)**, 2563-2570.

- Zhao, X., Yao, J., Liao, Z., Zhang, H., Chen, F., Wang, L., Lu, Y., Sun, X., Yu, S. and Tang, K. (2003). Molecular cloning of a novel mannose-binding lectin gene from *Arisaema heterophyllum*. *Plant Science* **165**, 55-60.
- Zhu, K., Bressan, R. A., Hasegawa, P. M. and Murdock, L. L. (1996). Identification of N-acetylglucosamine binding residues in *Griffonia simplicifolia* lectin II. *FEBS Letters* **390**, 271-274.

7. APPENDICES

APPENDIX A

Abbreviations for Amino Acids

Amino Acid	Three Letter	One Letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asparagine or Aspartic acid	Asx	B
Cysteine	Cys	C
Glycine	Gly	G
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glutamine or Glutamic acid	Glx	Z
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

APPENDIX B

Abbreviations for Lectins

BPA	<i>Bauhinia purpurea</i> agglutinin
Con A	concanavalin A from <i>Canavalia ensiformis</i>
DBL	<i>Dolichos biflorus</i> seed lectin
DB58	58 kDa vegetative lectin <i>Dolichos biflorus</i>
ECorL	<i>Erythrina corallodendron</i> lectin
FRIL	Flt3 receptor interacting lectin from <i>Dolichos lablab</i>
GNA	<i>Galanthus nivalis</i> agglutinin
GSi-B4	<i>Griffonia simplicifolia</i> isolectin B4
GS-II	<i>Griffonia simplicifolia</i> isolectin II
GS-IV	<i>Griffonia simplicifolia</i> isolectin IV
LCA	<i>Lens culinaris</i> agglutinin
LOLI	isolectin I from <i>Lathyrus ochrus</i>
MAH	<i>Maackia amurensis</i> haemagglutinin
MAL	<i>Maackia amurensis</i> leucoagglutinin
PAL	<i>Pterocarpus angolensis</i> lectin
PHA-E	erythroagglutinin from <i>Phaseolus vulgaris</i> (phytohaemagglutinin)
PHA-L	leucoagglutinin from <i>Phaseolus vulgaris</i> (phytohaemagglutinin)
PNA	peanut agglutinin (<i>Arachis hypogaea</i>)
PRA II	peanut root agglutinin II
PSL	<i>Pisum sativum</i> (pea) lectin
RPA	<i>Robinia pseudoacacia</i> (black locust) agglutinin
SBA	soybean agglutinin (<i>Glycine max</i>)
TxLC-1	tulip lectin
UEA I	isolectin I from <i>Ulex europeaus</i>
UEA II	isolectin II from <i>Ulex europeaus</i>
WBA-A	acidic winged bean agglutinin
WBA-B	basic winged bean agglutinin
WGA	wheat germ agglutinin
VVL-B4	<i>Vicia villosa</i> isolectin B4

APPENDIX C

Nucleotide and Amino Acid Sequence of Clone Muk151QII28

Muk151QII28	-----M L L N K A Y S -----ATGCTACTGAACAAAGCATACTC	23
1	Q D S L S F G F P T F P S D Q K	
Muk151QII28	CCAAGATTCCCTTTCTTCGGCTTCCCTACGTTTCCTTCAGACCAGAAAA Muk-23 and Muk-24	73
17	N L I F Q G D A Q I K N N A V Q L	
Muk151QII28	ACCTAATCTTCCAAGGTGATGCACAGATTAAAAACAACGCAGTCCAGCTT	123
34	T K T D S N G N P V A S T V G R I	
Muk151QII28	ACCAAGACAGACAGTAATGGTAACCCAGTGGCAAGCACTGTTGGGCGAAT	173
51	L F S A Q V H L W E K S S S R V	
Muk151QII28	CTTATTCTCGGCGCAAGTGCACCTTTGGGAAAAAAGCTCAAGCAGAGTGG	223
67	A N F Q S Q F S F S L K S P L S N	
Muk151QII28	CAAACCTTCAATCCCAGTTACAGCTTTTCCCTCAAATCACCGCTTTCCAAT	273
84	G A D G I A F F I A P P D T T I P	
Muk151QII28	GGAGCCGACGGCATTGCCTTCTTCATCGCACCGCCGGATACTACCATACC Muk-6	323
101	S G S G G G L L G L F A P G T A	
Muk151QII28	GAGCGGTTCTGGGCGGAGGTCTCCTAGGGCTCTTTGCACCTGGAAGTCTC AvrII	373
117	Q N T S A N Q V I A V E F D T F Y	
Muk151QII28	AAAACACGTCCGCGAACCAAGTGATTGCGGTTGAGTTTGACACCTTCTAT	423
134	A Q D S N T W D P N Y P H I G I D	
Muk151QII28	GCTCAAGACTCTAACACTTGGGATCCGAATTACCCGCACATTGGAATCGA ClaI	473
151	V N S I R S V K T V K W D R R D	
Muk151QII28	IGTCAACTCCATTAGGTCTGTGAAGACTGTGAAATGGGATAGGAGAGATG	523
167	G Q S L N V L V T F N P S T R N L	
Muk151QII28	GTCAATCCCTTAATGCCTCGTAACATTTAACCCCTAGCACTCGAAACCTG	573
184	D V V A T Y S D G T R Y E V S Y E	
Muk151QII28	GATGTCGTTGCTACTTACTCTGATGGTACAAGATACGAGGTGTCTTATGA	623
201	V D V R S V L P E W V R V G F S	
Muk151QII28	GGTTGACGTGAGGTCCGTGCTTCCAGAATGGGTTAGAGTTGGATTCTCAG	673
217	A A S G E Q Y Q T H T L E S W S F	
Muk151QII28	CTGCCTCAGGAGAGCAATATCAAACGCATACTCTTGAATCATGGTCTTTC	723
234	T S T L L Y T A Q K K G E N L A L	
Muk151QII28	ACCTCAACCTTGTGTACACTGCTCAGAAGAAGGGCGAGAATCTTGCCT Muk-25 Muk-12 TAGAATTCCC EcoRI	773
251	E M	
Muk151QII28	TGAGATGTGATCTGGAGTGATTCTATGTCATATCCATCATGTATGATAT	823

(Adapted from GenBank Accession Number AJ426056).

APPENDIX D

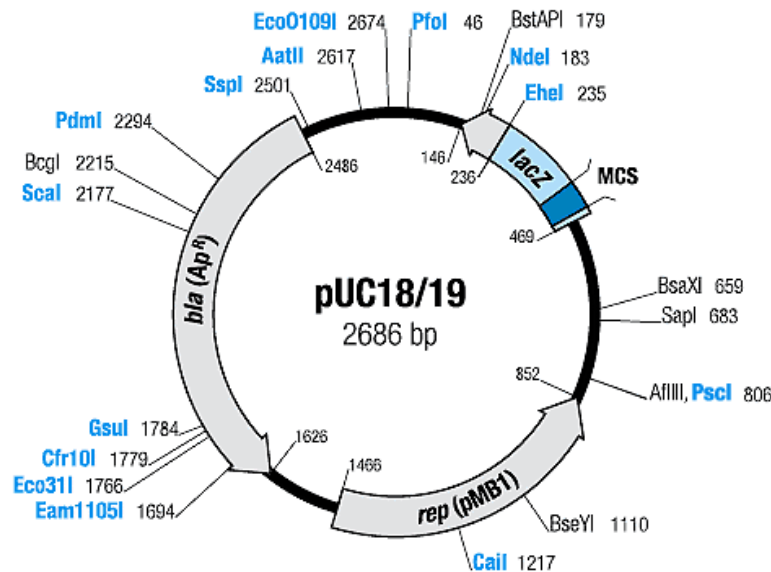
Primer Sequences

Primer	Size	Direction	Sequence (5'→3')
Muk-6	21-mer	→	CATCGCACCGCCGGATACTAC
Muk-12	19-mer	←	CTTCTTCTGAGCAGTGTAC
Muk-23	24-mer	→	CAAGATTCCCTTTCCTTCGGCTTC
Muk-24	25-mer	→	GATTCCCTTTCCTTCGGCTTCCCTA
Muk-25	34-mer	←	GGGAATT CTAAGCAGTGTACAACAAGGTTGAGGT
pBAD1	20-mer	→	CACACTTTGCTATGCCATAG
Muk-17	27-mer	→	CCTCAGGAGAGC GATAT CAAACGCATA
Muk-18	42-mer	→	GGGTTAGAGTTGGATTCT CCGGAG CCTCAGGAGAGCGATATC
Muk-19	27-mer	→	CCGGAGGCGTTGGCAACG CGGCCGAAT
Muk-20	23-mer	←	ATTCGGCCGCGTTGCCAACGCCT
Muk-21	30-mer	→	CCGGAGCG ACCGGT GCGCAGCGTGATGCGG
Muk-22	26-mer	←	CCGCATCACGCTGCGCACCCGGTCGCT
Muk-35	45-mer	→	GGAGCGACCG GTGCAC AGCGTGATGCGGCTCAAACGCATACTCTT
Muk-30	36-mer	→	CAGCTGCCTCAGGAG GATAT CAAACGCATACTCTTG
Muk-33	52-mer	→	GATTCTCAGCTGCCTCAGGAGGAG GATAT CAAACGCATACTCTTGAATCATG
Muk-34	45-mer	→	GACACCTTCTAT GCGCA AGACTATAACACTTGGGATCCGAATTAC
SELEC PUC1	25-mer	→	GAGTGCACCAT G GGCGGTGTGAAAT
SELEC PUC2	25-mer	→	GAGTGCACCAT ATG CGGTGTGAAAT
pUC FP	24-mer	→	CGCCAGGGTTTTCC CAGT CACGAC
pUC RP	23-mer	←	AGCGGATAACATTT CACAC AGGA

Restriction endonuclease recognition sites are shown in bold.

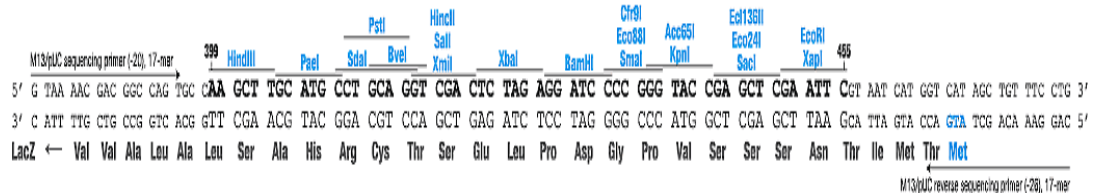
APPENDIX E

Plasmid Vector Map of pUC18/19

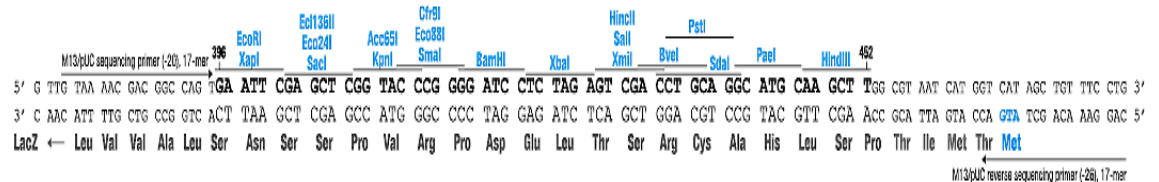


pUC18 and pUC19 are identical except that they contain the multi cloning sites (MCS) arranged in opposite directions. The positions of the enzymes that cut the plasmid once are shown. The rep(pMB1) represents a replicon derived from plasmid pBR322 and is responsible for replication of the plasmid. The *bla* gene, also derived from pBR322, codes for beta-lactamase that confers resistance to ampicillin. Alpha complementation is possible as the plasmid contains the α -peptide of the β -galactosidase (*lac Z*) gene.

MCS of pUC18



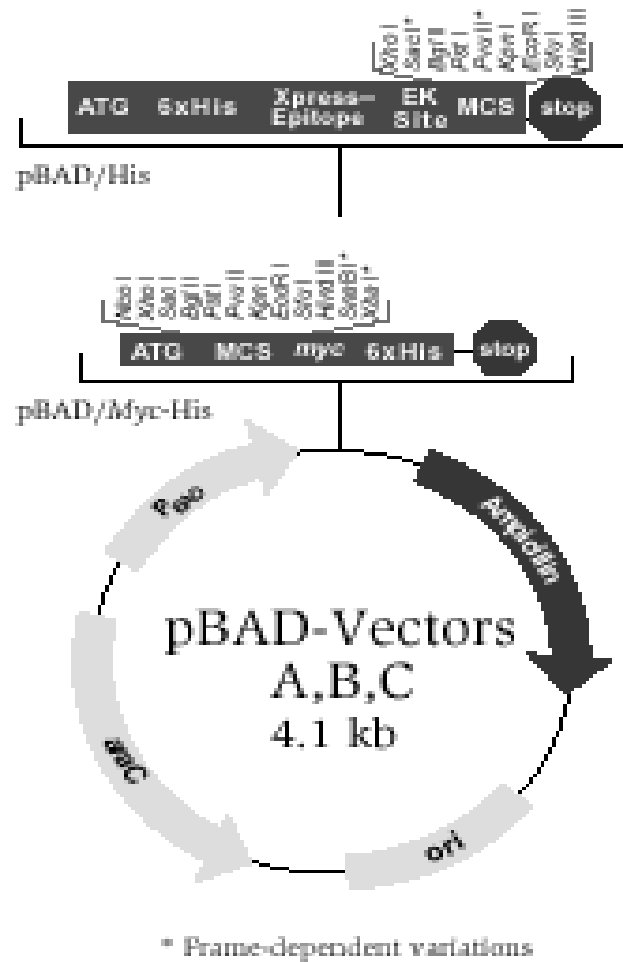
MCS of pUC19



(taken from www.fermentas.com/techinfo/nucleicacids/mappUC18/19.htm)

APPENDIX F

Plasmid Map of the Expression Vector pBADHis/pBADMycHis

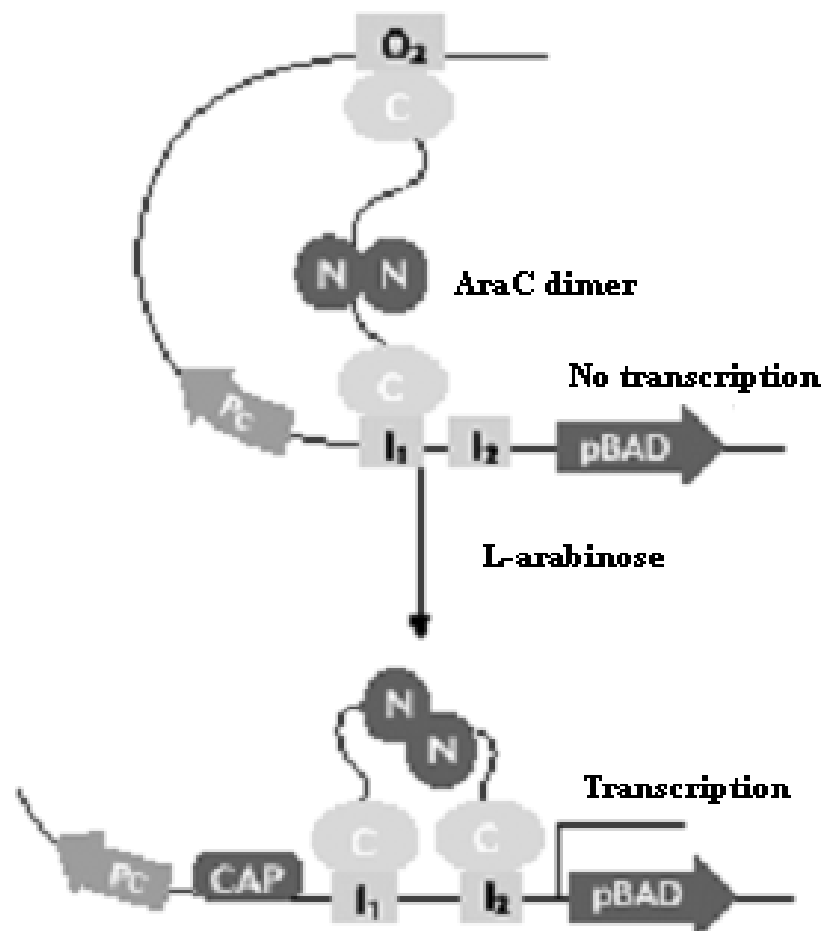


ATG represents the initiator codon; 6 x His, the polyhistidine tag for purification with ProBond resin; Xpress and *myc* epitopes for detection with antibodies directed against these epitopes; EK site, the enterokinase cleavage site and MCS, the multi-cloning site.

(taken from Invitrogen Life Technologies, 2000)

APPENDIX G

Regulation of the *araBAD* Promoter



The *araBAD* promoter is positively and negatively regulated by the product of the gene *araC*, AraC. AraC is a transcriptional regulator that forms a complex with arabinose. In the absence of arabinose, the AraC dimer binds O_2 and I_1 forming a 210 bp loop, a conformation that leads to the complete inhibition of transcription. In the presence of arabinose, the dimer is released from O_2 and binds I_1 and I_2 leading to transcription. The cAMP activator protein (CAP)-cAMP complex binds to the DNA and stimulates binding of AraC to I_1 and I_2 .

(taken from Invitrogen Life Technologies, 2000)

APPENDIX H

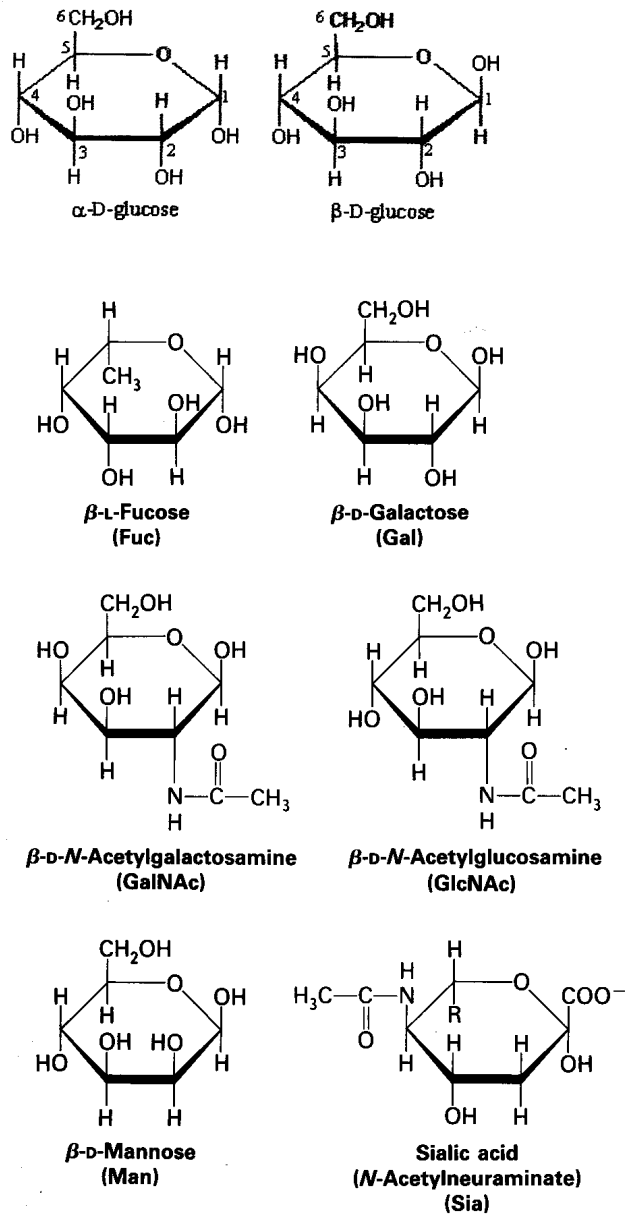
The Genetic Code

UUU UUC UUA UUG	phe leu	UCU UCC UCA UCG	ser	UAU UAC UAA UAG	tyr stop stop	UGU UGC UGA UGG	cys stop trp
CUU CUC CUA CUG	leu	CCU CCC CCA CCG	pro	CAU CAC CAA CAG	his gln	CGU CGC CGA CGG	arg
AUU AUC AUA AUG	ile met	ACU ACC ACA ACG	thr	AAU AAC AAA AAG	asn lys	AGU AGC AGA AGG	ser arg
GUU GUC GUA GUG	val	GCU GCC GCA GCG	ala	GAU GAC GAA GAG	asp glu	GGU GGC GGA GGG	gly

(Adapted from Devlin, 1992)

APPENDIX I

Structures of Selected Sugars



(taken from Stryer, 1988)