Introduction

It is just over a hundred years ago that the first plant lectin was described by Stillmark (Stillmark, 1888, 1889). Working with extracts from castor bean (*Ricinus communis*) he obtained a preparation which agglutinated red blood cells. As more and more of such substances were later discovered in other plants, and as a common name for them, the term, haemagglutinin, was proposed by Elfstrand (1898). The striking toxicity of some of the haemagglutinins, such as ricin and abrin (*Abrus precatorius*) has made the effects of these substances relatively easy to test on animals. Ehrlich, who is usually considered to be the father of immunology, has shown that rabbits fed with small amounts of seeds containing the toxin became partially immune to the toxicity, thus demonstrating that the haemagglutinins were also antigenic. Landsteiner and Raubitschek (1908) showed later that not all haemagglutinins need necessarily be as toxic as ricin or abrin. For example, the agglutinins obtained from common beans (*Phaseolus vulgaris*), peas (*Pisum sativum*), lentils (*Lens culinaris*), etc, were relatively non-toxic, water soluble proteins. It is now known that such haemagglutinating proteins are found in all taxonomic groups of the Plant Kingdom and that they are not all overtly toxic.

The next momentous step in the history of haemagglutinins was the realization that some of the haemagglutinins agglutinated blood cells only from some groups of individuals within the ABO blood group system, without affecting cells from other groups (Renkonnen, 1948; Boyd & Reguera, 1949). Indeed, this discovery of blood group specificity has led Boyd to coin the term, lectin, to denote this aspect of selection (in Latin, *legere* means to select) and is regarded as the starting point of modern lectinology.

With our increased understanding of the chemical structure of blood group-specific glycoconjugates, the involvement of sugars in the agglutination reaction was soon appreciated. As early as 1936, Sumner and Howell observed that cane sugar inhibited the haemagglutination activity of concanavalin A (Sumner & Howell, 1936). It was, however, somewhat later that Watkins and Morgan (1952) laid the foundations of our knowledge of the strict sugar specificity of the agglutination reaction within the human ABO blood group system. Indeed, the first proper definition of lectins was based on the sugar specificity of the inhibition of the haemagglutination reaction. Accordingly, lectins are carbohydrate-binding proteins of non-immune origin which agglutinate cells or precipitate polysaccharides or glycoconjugates (Goldstein *et al.*, 1980). This definition was adopted by the Nomenclature Committee of the International Union of Biochemistry (Dixon, 1981).

The main problem with this definition is that, if it is strictly interpreted, some poorly agglutinating well-known toxins, such as ricin, abrin, modeccin, etc, cannot be regarded as lectins, even though they are all known to contain lectinic subunits. Thus, the first definition has since been extended to include the above toxins (Kocourek & Horejsi, 1983). Moreover, as it is now realized that some lectins contain a second type of binding site that interacts with non-carbohydrate ligands (Barondes, 1988), confining the definition of lectins strictly to bivalent carbohydrate-binding proteins seems to have lost its usefulness. In fact, a narrow definition may even impede our understanding of the proper endogenous function(s) of lectins. Although the more general definition of lectins (Kocourek & Horejsi, 1983) is not yet universally accepted, it is very appealing and serves as a basis for the definition of lectins in this book. Accordingly, lectins are proteins (or glycoproteins) of non-immunoglobulin nature capable of specific recognition of, and reversible binding to, carbohydrate moieties of complex glycoconjugates without altering the covalent structure of any of the recognized glycosyl ligands. Thus, other sugar-binding proteins, such as the various sugar-specific enzymes, hormones and transport proteins are excluded, but monovalent lectins (i.e. bacterial and plant toxins) are included.

1

Lectins and their specificity

The carbohydrate specificity of lectins is established most conveniently by the Landsteiner hapten-inhibition technique (Landsteiner, 1962; Goldstein & Hayes, 1978; Goldstein & Poretz, 1986). In this method the inhibitory effectiveness of various mono- and oligosaccharides of known composition is compared in a recognized and convenient reaction of lectins, such as haemagglutination. Initially, on the basis of their reactivity with monosaccharides that differed in configuration at C-3 or C-4 of the pyranose ring, Mäkelä (1957) suggested that lectins could be divided into four classes. Thus, for example, concanavalin A, which reacts with D-mannose and/or D-glucose, belongs to group III. Soyabean lectin, whose specificity is for *N*-acetyl-D-galactosamine and/or D-galactose, is classified as group II. According to Mäkelä's scheme the L-fucose-binding lectins are members of group I (Fig. 1.1). However, no group IV-specific lectins have been found in Nature so far. In more recent studies, although still within the four basic classes of Mäkelä, the definition of the specificity of lectins has been further refined and extended. Accordingly, from their reaction with simple sugars, lectins are first classified into broad groups of lectin classes as either mannose/glucose-specific, or *N*-acetylglucosamine-specific, or *N*-acetylgalactosamine/galactose-specific or fucose-specific lectins. This is then followed by a more precise classification based on extensive investigations with a great number of oligosaccharides of known composition and structure. From such studies of the most complementary carbohydrate structure, the recognition and binding site of the lectins within the four main classes, is obtained.

Lectins are known to react chiefly with the non-reducing end of oligo- and polysaccharides. Although there are a number of known exceptions to this rule, the specificity of lectins for the terminal sugar tolerates little

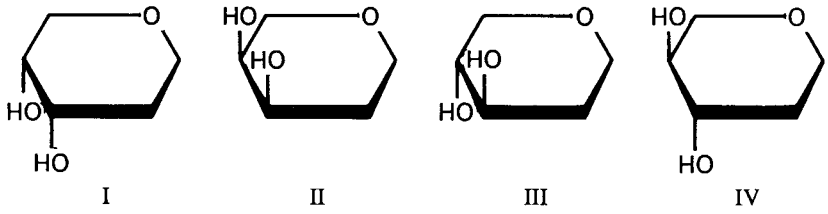

Fig. 1.1. Mäkelä's four major classes of lectins based on the configuration of hydroxyl groups at C-3 or C-4 of the pyranose ring.

variation at C-3 or C-4 of the pyranose ring. The configuration at the second carbon atom, however, appears to be less critical.

In view of their strict carbohydrate specificity, it is not surprising that most lectins have been purified to a high degree of purity by affinity methods. Thus, although some lectins, particularly those with unknown or unusual sugar specificities, in the past, have also been purified to homogeneity by the application of conventional protein purification methods, isolation based on affinity chromatography (Goldstein & Poretz, 1986) is now almost the rule. Indeed, in most instances, homogeneous lectin preparations can easily be obtained, in one-step procedures, by chromatography on affinity columns which contain insolubilized specific haptens bound to suitable supports.

D-mannose/D-glucose-specific lectins

This group contains the best known and most studied lectin: concanavalin A from jack bean (*Canavalia ensiformis*). Its structure, however, is somewhat atypical as concanavalin A is composed of four identical subunits. In contrast, most other lectins in the group, such as those obtained from peas (*Pisum sativum*), broad beans (*Vicia faba*) or lentils (*Lens culinaris*) consist of two light (α) and two heavy (β) chains and have a general subunit structure of $\alpha_2\beta_2$.

Purified preparations of concanavalin A were obtained as early as 1936. Indeed, this lectin was one of the earliest examples of a well-characterized crystalline protein described (Sumner & Howell, 1936). It was also one of the first lectins isolated by affinity chromatography (Agrawal & Goldstein, 1967). At or above pH 7, concanavalin A contains four non-glycosylated polypeptide chains of $M_r = 26{,}500$. Each subunit is made up of 237 amino acid residues. Although some of the peptide bonds in the primary sequence of the subunits are known to be broken, as the fragmented polypeptide chains are held together by non-covalent forces, the native conformation of the protein molecule is maintained regardless of whether the subunits are fragmented or not (Wang, Cunningham &

Edelman, 1971). The major breakage point in the polypeptide chain has been shown to be between Asn(118) and Ser(119) residues. When such concanavalin A preparations are dissolved in dissociating solvents, such as SDS, and electrophoresed on polyacrylamide gels in SDS, in addition to the major band of the unbroken subunit (M_r = 26,500), two smaller size polypeptide bands are also obtained. Rather interestingly, although concanavalin A is not a glycoprotein, it is, in fact, first synthesized as an inactive glycoprotein, with a leader sequence, in the developing seeds. In the inactive concanavalin A precursor, the amino terminal part of the active lectin which is characteristic for the lectin found in mature seeds, is in a midchain position. During seed development, the glycoprotein is modified by proteolytic processing and splitting at this midchain position and the leader sequence is removed from the amino terminus. The resulting two polypeptide chains are then ligated to each other in such a way that the amino terminal liberated from the middle of the precursor polypeptide chain becomes the new amino terminal of the mature form of the lectin, while the amino terminal of the precursor is joined to its carboxyl terminal residue by a new peptide bond. During this process, the carbohydrate part, which in the precursor shielded the carbohydrate-binding site of the lectin, is also removed. With this unmasking of its functional amino acid side-chains, the non-glycoprotein precursor is transformed into the fully active lectin, concanavalin A (Carrington, Auffret & Hanke, 1985) (Fig. 1.2).

Concanavalin A is a metalloprotein in which each subunit contains one Ca^{2+} and one Mn^{2+}. When the metal ions are removed, the lectin loses its activity (Agrawal & Goldstein, 1968). Apparently, the metal ions are necessary to lock the conformation of the lectin in a form in which the carbohydrate-binding sites are correctly exposed (Strazza & Sherry, 1982). Concanavalin A was also the first lectin whose complete three-dimensional structure has been determined by X-ray diffraction methods (Edelman *et al.*, 1972). Accordingly, the subunits of the lectin are compactly folded and dome-shaped structures. In the native concanavalin A, two of the subunit polypeptides with a structure based on antiparallel β-pleated sheets are joined in a functional ellipsoidal dimeric unit and two of these are then paired to form the native tetrameric lectin (Fig. 1.3).

The various functional sites in the lectin molecule have also been established. Thus, for example, the metal-binding sites are known to be in the amino terminal sequence of the lectin. The carbohydrate-binding site was determined on crystals obtained from the carbohydrate complexes of concanavalin A. Several amino acids are implicated in sugar

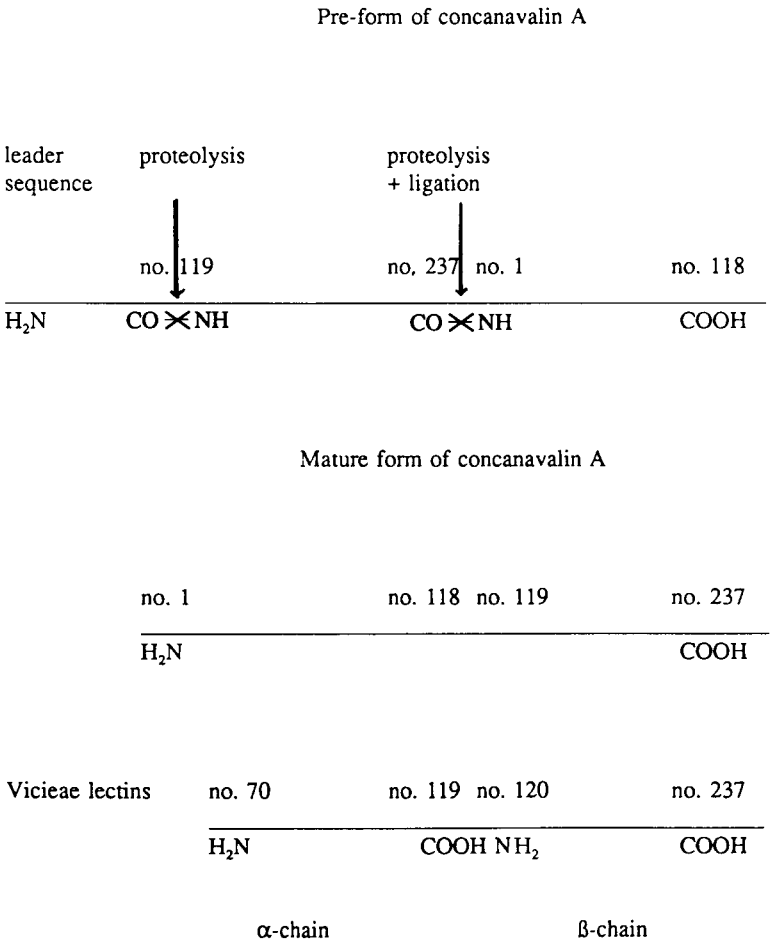

Fig. 1.2. Post-translational changes in the primary amino acid sequence of concanavalin A during seed maturation and homologies to other Vicieae lectins.

binding. Although these are in widely different positions in the primary amino acid sequence of the subunit polypeptides, spatially they are closely located in the native conformation of the lectin molecule. The amino acid residues involved are: Tyr(12), Asn(14), Asp(16), Leu(99), Tyr(100), Ser(168), Asp(208) and Arg(228).

In the classical early studies of Goldstein and his associates (for refs, see Goldstein & Poretz, 1986) it has been established that, of the monosaccharides, mannose, in its α-anomeric form, is the most active simple sugar inhibitor of the biological activity of concanavalin A. Moreover, in oligo- and polysaccharides, first and foremost, it is the non-

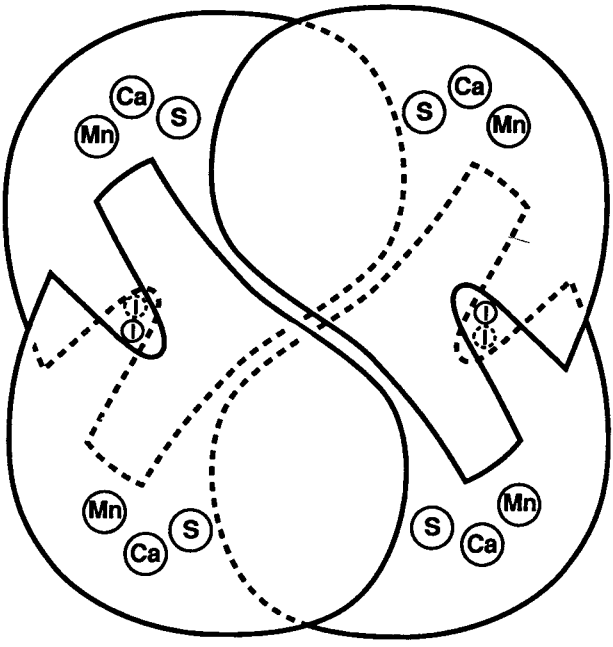

Fig. 1.3. Schematic representation of the tetrameric concanavalin A molecule with its metal (Ca^{2+} and Mn^{2+}), carbohydrate (S) and hydrophobic (I) binding sites (from Becker *et al.*, 1976; reproduced by permission of the authors and *Nature*).

reducing terminal mannose residue which is recognized by the lectin. Although this appears to be generally true for most polysaccharides interacting with concanavalin A, in α-(1–2) linked mannose-containing oligosaccharides, other mannose residues, and not just the terminal non-reducing ones, make appreciable contributions to the interaction. The results have, in fact, suggested that the best fit for the concanavalin A-carbohydrate-binding site is likely to be oligosaccharides containing three to four α-(1–2) mannosyl residues. Concanavalin A binds to branched trimannosyl structures, present in ovalbumin glycopeptides, with great affinity. These show similarity to the synthetic 3,6-di-*o*-(α-mannosyl)-mannose (Brewer *et al.*, 1985). Moreover, concanavalin A can be precipitated with this high mannose glycopeptide (D_3) from ovalbumin. Thus, one lectin reactive site, a trimannosyl moiety, is found on the α-(1–6) arm of the core β-mannose residue. Another site is found on the α-(1–3) arm of the core containing an α-(1–2) mannobiosyl group. Similar high mannose-type glycopeptides on the surface of a number of cells may also function as bivalent cross-linking agents and play a part in lectin-

induced patching and capping. Thus, these surface sugars may be involved in signal transduction as specific receptors (Bhattacharyya & Brewer, 1986).

Concanavalin A contains a number of non-polar binding sites of varying affinities. A high affinity site, which is close to the carbohydrate-binding site and reacts with phenyl groups in phenylglucosides and phenylmannosides, has been described (Poretz & Goldstein, 1971). Another high affinity site interacts with fluorescent hydrophobic probes (Roberts & Goldstein, 1983*a*) and one low affinity site per subunit binds tryptophan and indoleacetic acid (Edelman & Wang, 1978). There is also a non-polar binding cavity. This, however, is present only in concanavalin A crystals. It is clear that, with suitable ligands, non-polar binding sites can make considerable contributions to the primary sugar-specific binding and result in an increased and tighter overall binding between lectin and ligands. The functional importance of hydrophobic binding is clearly shown by the observation that even monomeric concanavalin A derivatives agglutinate guinea pig erythrocytes (Ishii *et al.*, 1984). The more recent demonstration of the specific binding of *myo*-inositol by concanavalin A, which is independent of the mannose/glucose binding site, further demonstrates that interactions between lectins and naturally occurring glycoconjugates are not necessarily confined to specific carbohydrate structures in the ligands (Wassef, Richardson & Alving, 1985). Thus, concanavalin A binds strongly to phosphatidylinositol-containing liposomes, even when these contain no sugar residues.

Several other lectins belonging to this class of mannose/glucose-specific haemagglutinins have been described in some detail over the years. One of these is the rather interesting close homologue of concanavalin A, the lectin isolated from the seeds of *Dioclea grandiflora* (Moreira *et al.*, 1983). Indeed, both plants from which these lectins have been isolated belong to the same tribe of Diocleae. The *Dioclea grandiflora* lectin is a metallo-protein, devoid of covalently bound carbohydrate residues and has the same specificity for D-mannose/D-glucose residues as concanavalin A. Its molecular weight and subunit structure, dissociation properties and the fragmentation patterns of its subunits resemble those of concanavalin A. Although 53 of the 237 amino acid residues of the intact α-chain polypeptide are different from those found in concanavalin A, six out of the total seven residues implicated in metal binding are conserved in the *Dioclea grandiflora* lectin (Richardson *et al.*, 1984).

Several mannose/glucose-binding lectins from the *Lathyrus* genus have been studied extensively. These lectins all appear to be very similar proteins containing no covalently bound carbohydrate residues, with M_r

values close to 50,000 and with a general subunit structure of $\alpha_2\beta_2$. Their specificity is similar to that of concanavalin A.

The class of D-mannose/D-glucose-binding lectins also contain a lectin isolated from lentils (*Lens culinaris*) with its main specificity for non-reducing α-mannopyranosyl terminal residues. This lectin also binds to oligo and polysaccharides which contain internal 2-*o*-substituted mannose residues. Glycoconjugates in which an α-L-fucosyl group is linked to *N*,*N*′-diacetylchitobiosyl units are also known to bind tightly to the lentil lectin. Similar lectins, with sugar-binding and molecular properties close to those of the lentil lectin, have been isolated and studied, from amongst others, peas (*Pisum sativum*), common vetch (*Vicia cracca*) and other members of the *Vicia* genus and sainfoin (*Onobrychis viciifolia*) (see Goldstein & Poretz, 1986).

N-acetyl-D-glucosamine-binding lectins

In this diverse group of lectins, whose members are found in Gramineae, Solanaceae and also in Leguminoseae, carbohydrate specificity is directed against *N*-acetyl-D-glucosamine and, particularly, its β-(1–4) linked oligomers.

The lectin from potato (*Solanum tuberosum*) tubers, a glycoprotein containing a fairly large covalently bound carbohydrate component, was first purified by conventional protein purification methods (Marinkovich, 1964). The lectin is unusual as it is readily inactivated by reduction with 2-mercaptoethanol. Purification of the lectin, however, is much more easily achieved by affinity chromatography. Indeed, several different affinity supports have been used successfully in the past for the isolation of homogeneous preparations of potato lectin. These, for example, have included fetuin-Sepharose (Owens & Northcote, 1980) or *N*,*N*′,*N*″-triacetylchitotriose–Sepharose (Desai & Allen, 1979).

Potato lectin has a molecular weight of about 100,000 and contains two non-covalently linked identical subunits (Allen & Neuberger, 1973; Owens & Northcote, 1980). Each subunit of the glycoprotein lectin is made up of two dissimilar domains. One of the domains contains all the hydroxyproline and carbohydrate residues. This part of the molecule also contains about half of the total number of serine residues and these are all galactosylated. The carbohydrate part of this domain contains mainly L-arabinose. Indeed, over 90% of the sugar constituents of the carbohydrate part consists of arabinose, while the rest is galactose (Allen *et al.*, 1978). In fact, the overall composition of this domain of the potato lectin resembles closely most of the structural features of plant cell wall

glycoconjugates. The bulk of the arabinose occurs in short linear chains of three or four arabinofuranosyl groups linked to hydroxyproline (Allen *et al.*, 1978; Ashford *et al.*, 1982). All linkages in the triarabinoside are β-(1–2) and in the tetraarabinoside the last residue is linked by an α-(1–3) linkage (Ashford *et al.*, 1982).

The other domain of the lectin subunit is very different and contains all the cysteine residues of the molecule. The carbohydrate-binding site of the lectin is also located in this domain. Not surprisingly, due to the substantial differences in the two domains, antibodies raised against the intact potato lectin contain two totally different populations of immunoglobulins. One population reacts exclusively with the glycosylated part of the molecule, while the cysteine-rich part of the lectin is recognized only by the second immunoglobulin component of the antibody mixture. The fundamental difference between the two domains is also shown by the observation that antibodies against the glycosylated domain cannot abolish the biological activity of potato lectin, while reaction of the intact lectin with the antibody population against the cysteine-rich part strongly inhibits haemagglutination activity. Similarly, most of the covalently bound sugar residues can be removed from the lectin by treatment with trifluoromethanesulphonic acid, without affecting its activity. Accordingly, the active site of potato lectin containing the carbohydrate recognition and binding sites is confined to the cysteine-rich and non-glycosylated domain of the lectin molecule.

Oligosaccharides of *N*-acetyl-D-glucosamine are all good inhibitors of the haemagglutination activity of potato lectin (Allen & Neuberger, 1973). However, *N*-acetyl-D-glucosamine itself is very poorly inhibitory. In fact, the inhibitory potency of the monosaccharide is of six orders of magnitude less than that of *N*,*N*′-diacetylchitobiose. Moreover, in *N*-peracetylated chitooligosaccharides, the inhibitory activity increases with increasing chain length. Thus, *N*,*N*′,*N*″,*N*‴,*N*⁗-pentaacetylchitopentaose is about 50 times more effective in inhibiting the haemagglutination activity of potato lectin than *N*,*N*′-diacetylchitobiose. These results indicate that the sugar-complementary site of potato lectin is fairly extensive in size. With equilibrium dialysis technique Ashford *et al.* (1982) showed that each subunit (M_r = 50,000) has only one reactive site. Thus, for high haemagglutination potency of the native dimer lectin, both subunits take part in carbohydrate binding.

A lectin, similar to the extensively studied tuber lectin, has also been isolated from the pericarp of the potato fruit (Kilpatrick, 1980*b*). This lectin is very similar to the tuber lectin in both molecular properties and activity. In fact, all lectins from potatoes (and even from *Datura* species)