ELSEVIER

Biochimica et Biophysica Acta 1383 (1998) 9–36

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Review

Legume lectin structure

Remy Loris *, Thomas Hamelryck, Julie Bouckaert, Lode Wyns

Laboratorium voor Ultrastruktuur, Vlaams Interuniversitair Instituut voor Biotechnologie, Vrije Universiteit Brussel, Paardenstraat 65, B-1640 Sint-Genesius-Rode, Belgium

Received 16 July 1997; revised 13 October 1997; accepted 16 October 1997

Abstract

The legume lectins are a large family of homologous carbohydrate binding proteins that are found mainly in the seeds of most legume plants. Despite their strong similarity on the level of their amino acid sequences and tertiary structures, their carbohydrate specificities and quaternary structures vary widely. In this review we will focus on the structural features of legume lectins and their complexes with carbohydrates. These will be discussed in the light of recent mutagenesis results when appropriate. Monosaccharide specificity seems to be achieved by the use of a conserved core of residues that hydrogen bond to the sugar, and a variable loop that determines the exact shape of the monosaccharide binding site. The higher affinity for particular oligosaccharides and monosaccharides containing a hydrophobic aglycon results mainly from a few distinct subsites next to the monosaccharide binding site. These subsites consist of a small number of variable residues and are found in both the mannose and galactose specificity groups. The quaternary structures of these proteins form the basis of a higher level of specificity, where the spacing between individual epitopes of multivalent carbohydrates becomes important. This results in homogeneous cross-linked lattices even in mixed precipitation systems, and is of relevance for their effects on the biological activities of cells such as mitogenic responses. Quaternary structure is also thought to play an important role in the high affinity interaction between some legume lectins and adenine and a series of adenine-derived plant hormones. The molecular basis of the variation in quaternary structure in this group of proteins is poorly understood.

Keywords: Lectin; Protein–carbohydrate interactions; Quaternary structure; Carbohydrate recognition

Contents

Abbreviations: PHA-L, phytohemagglutinin-L from *Phaseolus vulgaris*; PHA-E, phytohemagglutinin-E from *Phaseolus vulgaris*; SBA, soybean agglutinin; Con A, concanavalin A; GS-IV, *Griffonia simplicifolia* lectin IV; GS-II, *Griffonia simplicifolia* lectin II; LOL, *Lathyris ochrus* lectin; LOL I, *Lathyris ochrus* isolectin I; LOL II, *Lathyris ochrus* isolectin II; PNA, peanut agglutinin; DBL, *Dolichos biflorus* seed lectin; UEA-I, *Ulex europaeus* lectin I; UEA-II, *Ulex europaeus* lectin II; MDP, *N*-acetylmuramyl dipeptide; NMR, nuclear magnetic resonance; NMRD, nuclear magnetic resonance dispertion

* Corresponding author. Fax: 0032-2-3590289; E-mail: reloris@vub.ac.be remy@ultr.vub.ac.be

0167-4838/98/$19.00
PII S0167-4838(97)00182-9

1. Introduction

Recognition between proteins and carbohydrates is of prime importance in many biological processes, such as viral, bacterial, mycoplasmal and parasitic infections, targeting of cells and soluble components, fertilisation, cancer metastasis and growth and differentiation. Legume lectins are the model system of choice to study the molecular basis of these recognition events because they are not only easy to purify in large quantities, but also exhibit a wide variety of carbohydrate specificities despite strong sequence conservation [1].

Legume lectins have, for decades, been a paradigm in the area of protein–carbohydrate recognition. Lectin activity in plant seeds has been demonstrated as early as 1888 [2]. Despite the lack of knowledge on their in vivo activity, they have proven to be useful tools in immunology and glycobiology because of their wide range of specificities for complex carbohydrates. Their carbohydrate specificity has also allowed them to be used in such applications as purification and characterisation of complex carbohydrates and glycoconjugates and bone marrow transplantation. Thus, the in vitro function of these proteins (carbohydrate recognition) is well documented and the molecular basis of this interaction has been studied with a variety of biophysical techniques, including X-ray crystallography, NMR and microcalorimetry.

Legume lectins are not only of interest because of their carbohydrate binding properties. Related to their potential to agglutinate cells and to precipitate multivalent carbohydrates is the fact that these proteins are oligomeric, forming either dimers or tetramers. Recent structural data have demonstrated that an essentially strongly conserved monomeric unit can oligomerise in a variety of ways, making these proteins also of interest to analyse the details of protein–protein interaction. Furthermore, quaternary structure relates to activity, as has been shown in recent years for a variety of lectins: highly ordered homogeneous cross-linked lattices are formed when lectins are mixed with multivalent carbohydrates, resulting in a higher form of specificity than can be achieved on the level of the monomer. In this review, we will focus on the structure of legume lectins and analyse the structural basis of their carbohydrate specificity.

2. The legume lectin monomer

The legume lectins form a large family of homologous proteins. Of about 50 legume lectin sequences that have been determined all show pairwise sequence identities not lower than 35%. Recently, it has been suggested that also in the animal kingdom legume lectin homologues may be present [3–5]. For 11 members of the legume lectin family, crystallo-

graphic co-ordinates have been deposited at the protein data bank (see also Tables 1 and 2). These include concanavalin A [6,7], PHA-L [8], the lectins from pea [9], lentil [10,11], *Lathyrus ochrus* [12], peanut [13,14], soybean [15], coral tree [16], *Griffonia simplicifolia* lectin IV [17]. The structures of the lectin-related α-amylase inhibitor from *Phaseolus vulgaris* [18] and arcelin [19], which have no known carbohydrate binding activity, are also known. The best refined structure to date is that of concanavalin A, which has been refined at 1.2 Å [20] and presently being further refined against 0.94 Å data [21]. Not surprisingly, the legume lectin monomer is structurally well conserved. It consists of two large β-pleated sheets that form a scaffold on which the carbohydrate binding region is grafted (Fig. 1(a)). The same architecture and topology is found in a wide variety of carbohydrate recognising proteins such as the galactins [22,23], serum amyloid protein [24], the lectin-like domains Wing-1 and Wing-2 attached to the catalytic domain of *Vibrio cholerae* neuraminidase [25] and Charcot–Leyden crystal protein [26]. The topology of the legume lectin fold, which is shown in Fig. 1(b), is complex and is structurally related to the jelly-roll topology commonly found in viral coat proteins [27,28]. The architecture of the legume lectin monomer is usually described as consisting of two β-sheets. As was pointed out by Banerjee and co-workers, the correct description involves three β-sheets: the 6-stranded back sheet, the 7-stranded front sheet and a smaller 5-stranded sheet (which we will call the S-sheet), that plays a major role in holding the two large sheets together [14]. The main hydrophobic core is located between the back and the front sheet. No α-helix is present and about 50% of the residues are in loop regions. One of these loops (an Ω-loop containing the conserved glycine of the monosaccharide binding site – see below) curls over the front sheet, resulting in the formation of a second hydrophobic core between the front sheet and this loop.

Carbohydrate binding activity of legume lectins depends on the simultaneous presence of both a calcium and a transition metal ion. These metal binding sites were first described in detail for concanavalin A [29] and have been found to be extremely well conserved in all other legume lectin structures. A schematic drawing of the metal binding sites in concanavalin A is given in Fig. 2(b). The calcium and the transition metal ion are approximately 4.5 Å apart and bridged by two aspartate residues. Both metals have four protein ligands and two water ligands. One of the water ligands from the calcium ion forms a bridge with the carbonyl group of an aspartate residue (Asp208 in concanavalin A) that is preceded by a *cis*-peptide bond.

Table 1
Crystal structures of uncomplexed legume lectins

Species	Comments	Resolution	R-factor	R_{free}-factor	PDB code
Canavalia ensiformis	Containing both a calcium and a transition metal ion	2.0–1.2	0.142–0.178	nr [a]	1JBC 1CON 1ENR 1SCS 1SCR 2CTV
Canavalia ensiformis	Triclinic form	2.4	0.205	0.265	1VLN
Canavalia ensiformis	Containing only a transition metal ion	2.5–2.8	0.198–0.207	0.249–0.284	1CES 1ENS 1ENQ
Canavalia ensiformis	Demetallised	2.5	0.180	0.247 [b]	1APN
Canavalia brasiliensis		3.0	0.163	0.246	ND [c]
Lathyrus ochrus I.		1.9	0.185	nr [a]	1LOE
Pisum sativum		1.7	0.177	nr [a]	2LTN
Lens culinaris	2 crystal forms	1.75/1.8	0.175/0.184	nr [a]	1LEN 2LAL
Phaseolus vulgaris	PHA-L	2.8	0.200	0.226 [b]	1FAT
Phaseolus vulgaris	Arcelin	2.7	0.206	0.271	1IOA
Phaseolus vulgaris	α-amylase inhibitor complexed with PPA	1.85	0.183	0.22	1DHK
Griffonia simplicifolia	Lectin IV	2.0	0.187	nr [a]	1LEC

[a] nr: not reported.
[b] a-posteriori determined R_{free}.
[c] ND: not deposited at the time of writing.

Table 2
Crystal structures of legume lectin–carbohydrate complexes

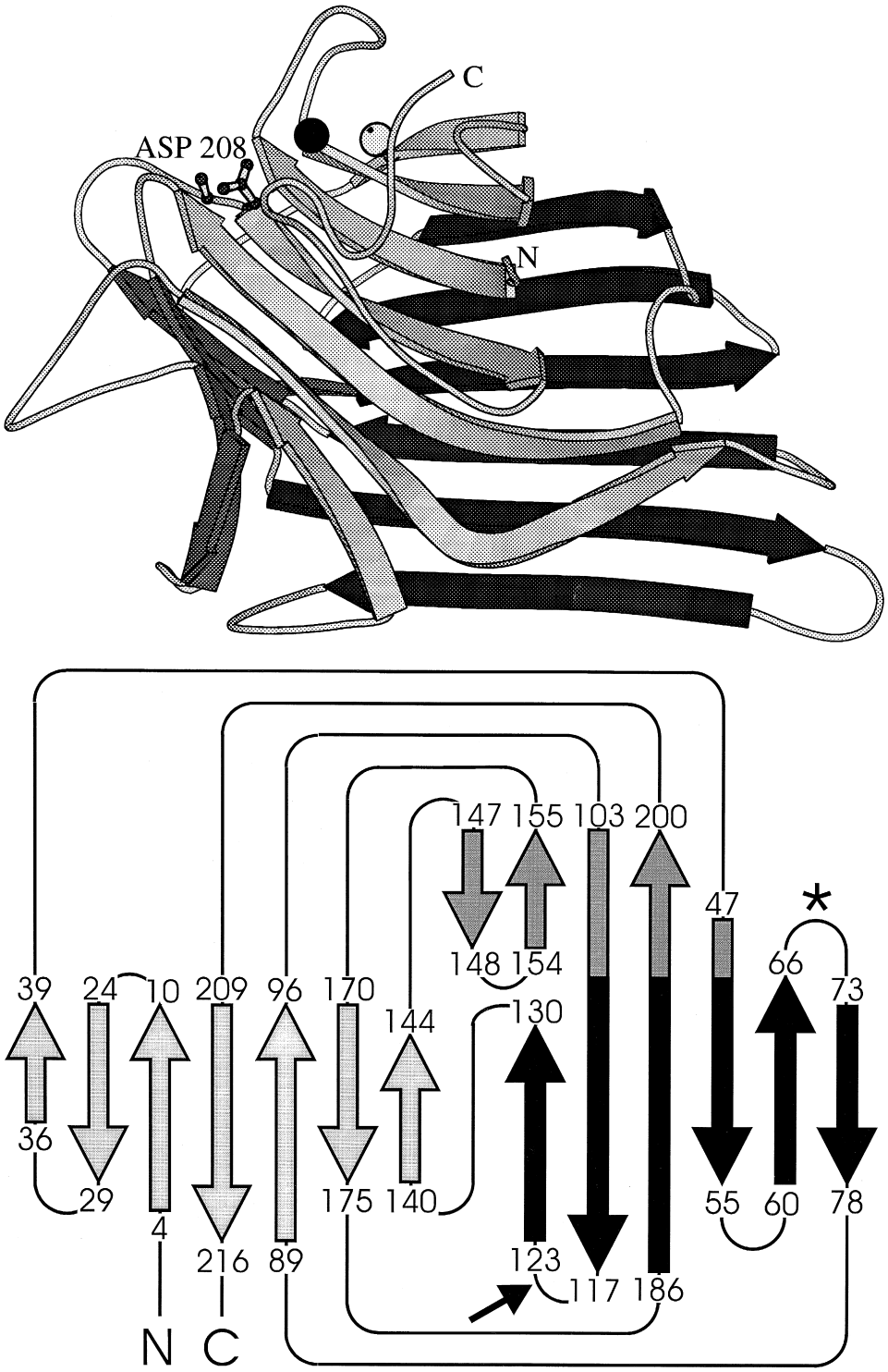
Species	Bound carbohydrate	Resolution	R-factor	R_{free}-factor	PDB code
Canavalia ensiformis	Methyl α-D-Mannopyranoside	2.0	0.199	nr [a]	5CNA
Canavalia ensiformis	Methyl α-D-Glucopyranoside	2.0	unrefined	nr [a]	ND [b]
Canavalia ensiformis	4′-Nitrophenyl-α-D-Mannopyranoside	2.75	0.185	0.260	1VAM
Canavalia ensiformis	4′-Nitrophenyl-α-D-Glucopyranoside	3.0	0.186	0.274	1VAL
Canavalia ensiformis	4′-Methylumbelliferyl-α-D-Glucopyranoside	2.78	0.182	0.216	1CJP
Canavalia ensiformis	Arabinose	2.7	unrefined	nr [a]	ND [b]
Canavalia ensiformis	Manα1–3[Manα1–6]Manα1-Me	2.3/2.4	0.205/o.201	0.255/0.255	1CVN/1ONA
Vicia faba	Glucose	2.8	0.37	nr [a]	ND [b]
Pisum sativum	Manα1–3[Manα1–6]Manα1-Me	2.6	0.183	nr [a]	1RIN
Lathyrus ochrus I	Methyl α-D-Mannopyranoside	1.96	0.182	nr [a]	1LOB
Lathyrus ochrus I	Methyl α-D-Glucopyranoside	2.2	0.179	nr [a]	1LOA
Lathyrus ochrus I	Muramic acid	2.05	0.197	nr [a]	1LOD
Lathyrus ochrus I	Mur-D-Ala-D-iGln (MDP)	2.05	0.189	nr [a]	1LOC
Lathyrus ochrus I	Manα1–3Manβ1–4GlcNAc	2.1	0.175	nr [a]	1LOG
Lathyrus ochrus I	Biantennary octasaccharide	2.3	0.190	nr [a]	1LOF
Lathyrus ochrus II	Biantennary glycopeptide from human lactotransferrin	2.8	0.185	nr [a]	1LGC
Lathyrus ochrus II	N2 fragment of human lactotransferrin	3.3	0.210	nr [a]	1LGB
Lens culinaris	Glucose	3.0	0.206	nr [a]	1LEM
Lens culinaris	Sucrose	1.9	0.188	nr [a]	1LES
Erythrina corallodendron	Galβ1–4Glc (Lactose)	2.0	0.190	nr [a]	1LTE
Griffonia simplicifolia IV	Fucα1–2Galβ1–3[Fucα1–4]GlcNAc (Lewis b)	2.0	0.181	nr [a]	1LED
Griffonia simplicifolia IV	Fucα1–2Galβ1–4[Fucα1–3]GlcNAc (Lewis Y)	2.0	0.185	0.242	1LED
Arachis hypogaea	Galβ1–4Glc (Lactose)	2.2	0.164	nr [a]	2PEL
Arachis hypogaea	Galβ1–3GalNAc (T-antigen disaccharide)	2.5	0.175	0.251	1TEP
Glycine max	(Galβ1–4GlcNAcβ[Galβ1–4GlcNAcβ]-Galβ-O$(CH_2)_5COOCH_3$	2.6	0.201	nr [a]	1SBA

[a] nr: not reported.
[b] ND: not deposited at the time of writing.

Recently, the structure of demetallised concanavalin A has been refined at 2.5 Å resolution and the structures of several intermediate steps in the remetallisation process have been presented [30,31]. As predicted by earlier NMRD measurements [32], the key event in the metal-induced activation of concanavalin A is the *trans* to *cis* isomerisation of an Ala–Asp peptide bond. The metal binding region in concanavalin A is partly unfolded upon demetallisation, and the conformations of several loops in the demetallised structure seem to depend on packing contacts. No calcium binding site is present in demetallised concanavalin A, but a proto-transition metal binding site is formed by the side chains of residues Glu8, Asp10 and His24 (Fig. 2(b)). This metal-free conformation is called the 'unlocked' form of ConA,

Fig. 1. The legume lectin monomer. (a) Ribbon diagram of the concanavalin A monomer. The *cis*-aspartate Asp208 is labeled and shown in ball-and-stick, while the manganese and calcium ions are indicated by grey and black spheres. The 6-stranded back sheet is coloured black, the 7-stranded front sheet light grey and the small S-sheet in dark grey. (b) Topology diagram of the legume lectin fold. Colour-coding of the strands is identical as in (a). The N- and C-termini are those found in concanavalin A. The position of these termini in the circularly permuted structures of most other legume lectins are indicated by an arrow. The cleavage site between the β-chain and the α-chain of te *Vicieae* lectins is indicated by an asterix. All figures (except Fig. 1(b) and Fig. 2(a), Fig. 6Fig. 10Fig. 11) were drawn with MOLSCRIPT [156].

while the active, metal-bound conformation is called the 'locked' form. Binding of zinc or cobalt in this proto-site initially leads to the movement of Asp19 and to an ordering of the loop Pro13–Pro23, which is partly disordered in the structure of apo-concanavalin A, but overall the protein stays in its unlocked apo conformation. Solution NMRD studies showed that about 10% of the molecules in demetallised concana-



valin A are in the locked conformation (capable of carbohydrate recognition) and that this percentage can increase significantly in the presence of high concentrations of mannose [33,34]. Thus, locked and unlocked conformations are in equilibrium and calcium (but also to a lesser extent manganese and cadmium) stabilises the locked conformation [35,36]. Indeed, presence of a metal bound in the calcium binding site is essential for observing a native conformation for the loops Pro13–Pro23 and Thr97–Glu102 of the metal and carbohydrate binding region and for observing a *cis*-peptide bond between Ala207 and Asp208 in crystal structures of concanavalin A. The all-or-nothing conformational transition observed in the remetallisation process of concanavalin A thus resembles to a certain extent protein folding, in the sense that one goes from an essentially unfolded and flexible metal binding region in the apo structure to a unique well ordered structure in the metal-bound form of the protein. The time constants in the concanavalin A system are, however, much larger than those observed for the folding of small globular proteins, due to the large energy barrier provided by the requirement of a *cis–trans* isomerisation step of an Ala–Asp peptide bond.

In this context, it is noteworthy that binding of monosaccharides to concanavalin A stabilises its native conformation and protects against chaotropic agents such as urea [146]. Since high concentrations of mannose are also capable of shifting the locked–unlocked equilibrium towards the locked conformation in the absence of metal ions [34], this stabilisation can be explained. The unlocked structure can be considered as partly unfolded, and saccharide binding therefore stabilises a conformation that has a higher degree of folding.

The legume lectin family also contains two proteins with no known carbohydrate recognition activity that do not possess the otherwise strictly conserved metal binding sites: the α-amylase inhibitor (α-AI) from *Phaseolus vulgaris* [37,145] and arcelin, a seed defence protein found in certain wild acces-

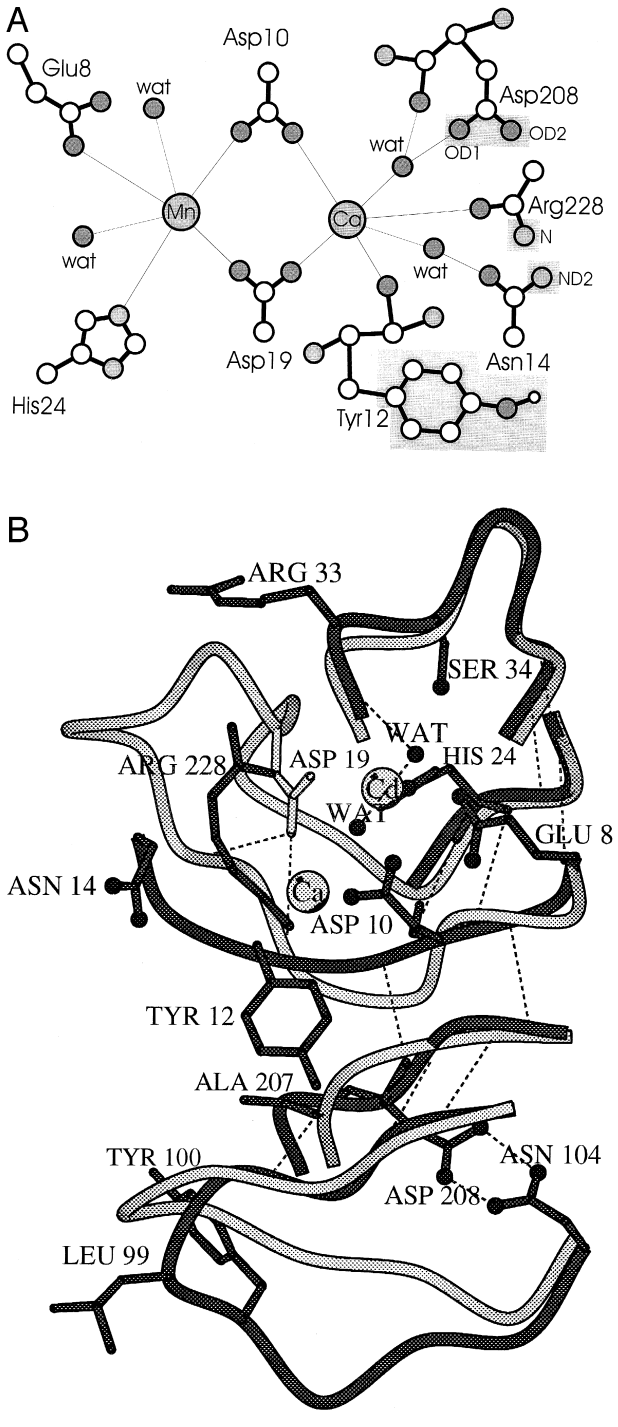


Fig. 2. (a) Schematic representation of the double metal binding site of concanavalin A. Carbon atoms are shown in white, nitrogen in light grey and oxygen in dark grey. The two metal ions are shown as larger light grey circles and are labeled. Functional groups that are esential for monosaccharide binding are drawn on a light-grey shaded background. Oxygen and nitrogen atoms from the conserved Asp-Asn–Arg(Gly) triad that hydrogen bond to a bound monosaccharide are labeled. Only side chain atoms are shown, except for Tyr12, Asp208 and Arg228. In the case of Arg228, side chain atoms were omitted for the sake of clarity. (b) Superposition of the metal binding region in holo (grey) and apo concanavalin A (black).

Table 3
Residues involved in metal binding in two typical legume lectins (concanavalin A and PHA-L) and in α-AI and arcelin

	Con A	PHA-L	Arcelin	α-AI
Pure transition metal ligands	Glu8	Glu122	Val123	Glu101
	His24	His137	Arg130	Arg108
Ligands bridging both metals	Asp10	Asp124	Asn125	Asp103
	Asp19	Asp132	–	–
Pure calcium ligands	Tyr12	Leu126	–	–
Calcium ligands involved in carbohydrate recognition	Asn14	Asn128	–	–
Carbohydrate ligands bridged to calcium via a water	Asp208	Asp86	Tyr85	Val80
	Arg228	Gly104	Arg103	–

sions of the same plant [38]. Both proteins show about 50–60% sequence identity to the two lectins from *Phaseolus vulgaris* PHA-L and PHA-E. The crystal structures of arcelin variant 5a and α-AI variant 1, as well as that of PHA-L have recently been solved [8,18,19]. Both α-AI and arcelin lack the loop corresponding to residues Pro13 to Pro23 in concanavalin A. Also, most of the other residues that in other legume lectins play a role in metal and carbohydrate binding are not conserved (Table 3 and

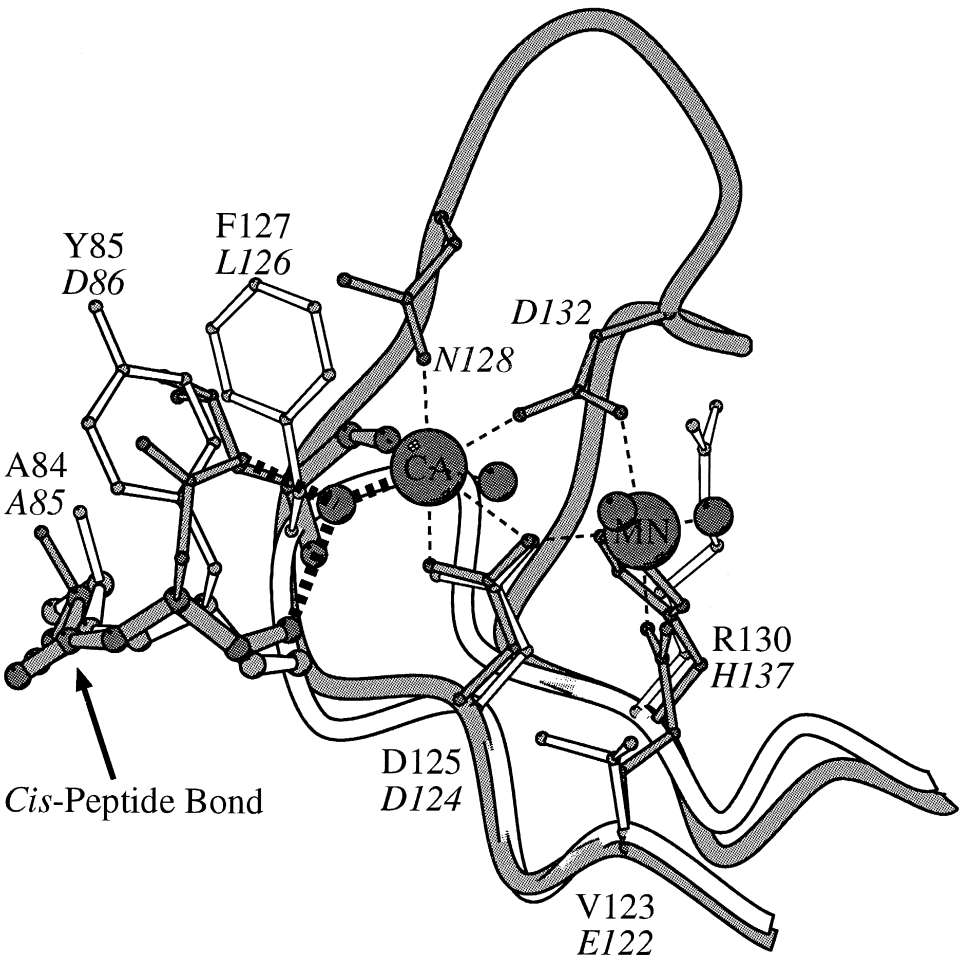


Fig. 3. Superposition of the metal binding region of PHA-L (grey) on the corresponding structure or arcelin (white) from the same plant (*Phaseolus vulgaris*). PHA-L residue are labeled normal, arcelin residues in italic.

Fig. 3). No bound metals are observed in the crystal structure of arcelin or α-AI. Nevertheless, in arcelin Tyr85, the residue corresponding to the *cis*-aspartate in the other legume lectins, is also preceded by a *cis* peptide bond [19]. This contrasts with the results obtained on demetallised concanavalin A and suggests that calcium binding is one way, but not the only way in which this particular *cis* peptide bond can be stabilised. Stabilising interactions include stacking between the aromatic rings of Tyr85 and Phe127 and two main chain – main chain hydrogen bonds (Ala84 NH to Thr205 O and Gly207 NH to Ala84 O). The presence of this unusual and unpredicted *cis*-peptide was interpreted in terms of the possible involvement of Tyr85 in an active role in the defence function of this protein.

The lectin-related α-amylase inhibitor from *Phaseolus vulgaris* (α-AI) shows an even larger truncation of the loops involved in metal- and monosaccharide binding by ordinary legume lectins [37,145]. In the recent structure of α-AI in complex with porcine pancreatic α-amylase, no *cis*-peptide is observed, although the binding site for porcine α-amylase is located in the same region of the protein as the carbohydrate binding sites in the lectins and the putative active site in arcelin [18].

3. Quaternary structure

3.1. The canonical legume lectin dimer

Lectins were originally defined as carbohydrate binding proteins from non-immune origin that are able to agglutinate cells or to precipitate complex carbohydrates, without having any enzymatic activity towards their carbohydrate ligands [39,40]. As a consequence, legume lectins have multiple binding sites and indeed, they are found to form dimers and tetramers. The only known exception to this rule in the legume lectin family is arcelin 5a, the lectin-related defence protein from *Phaseolus vulgaris*, which has no known carbohydrate binding potential. Although originally identified as a dimer at pH 7.0, its recently determined crystal structure showed a monomer.

Most known legume lectins contain a structure termed the ‘canonical legume lectin dimer’. This dimer, which was first described for concanavalin A and pea lectin, is characterised by a large 12-stranded β-sheet resulting from the association of the two 6-stranded back sheets (Fig. 4) [6,7,9]. Further contacts occur through side chain–side chain and side chain–main chain interactions of residues 14β to 17β, the hydrated β-bulge from residues 46β to 49β in the front sheet and residues 90β and 18α (pea lectin numbering).

In pea lectin, lentil lectin as well as *Lathyrus* lectin, the dimer interface is quite hydrophyllic and contains a number of buried water molecules [41]. Within a single species, the positions of these water molecules are well conserved, as is evident from an analysis of different crystal forms of lentil lectin and *Lathyrus* lectin. However, these waters are only poorly conserved among lectins from different species and are modulated by point mutations at the interface, and also by disturbances from perfect 2-fold symmetry by strong packing interactions, as demonstrated by the crystal structure of pea lectin.

3.2. Other dimers

Two dimeric lectins of known structure do not form the canonical legume lectin dimer. These are the lectins from coral tree (*Erythrina corallodendron*) [16] and lectin IV from *Griffonia simplicifolia* [17]. These structures are compared with the canonical dimer in Fig. 4. In the case of the *E. corallodendron* lectin, it was suggested that the canonical dimer cannot be formed due to sterical hindrance from a carbohydrate covalently bound to Asn17 [16]. An alternative dimer is formed with a buried surface of only 700 Å^2 per monomer, compared to the about 1100 Å^2 buried in the interface of the canonical dimer.

GS-IV forms a heterodimer, with both monomers differing in their covalently bound carbohydrates: Asn5 is glycosylated in only one of the two while Asn18 is glycosylated in both [17]. Inter-subunit interactions consist of side chain–side chain and side chain–main chain interactions as well as water bridges between residues Ser71, Thr159–Arg160, Thr177–Thr186, Ile198–Ser200 and Asp239, with a buried surface area of 900 Å^2 per subunit. Formation of the canonical dimer is thought not to be impaired by the glycosylation state. Instead, formation of the standard dimer may be prevented because this would bury a charged residue (Glu58) in the subunit interface. The corresponding residue in the lentil lectin dimer is Thr48β and is located close to the local dyad axis

relating the two subunits in the canonical dimer. In the lentil lectin dimer, the Thr48 side chains from both subunits hydrogen bond to each other, while in the homologous pea and *Lathyrus* lectin interfaces the corresponding Ser48β are differently oriented, resulting in a different solvent network in the interfaces of these two proteins. Formation of a canonical GS-IV dimer would thus bring two negatively charged groups within 4 Å from each other. Furthermore, in a canonical dimer environment, the side chain of Glu58 would clash with the side chain of Tyr56 of the other monomer.

3.3. The concanavalin A tetramer

The quaternary structure of concanavalin A is pH dependent, the protein being tetrameric above pH 7.0 and dimeric below pH 5.0 [42]. The dimer formed at

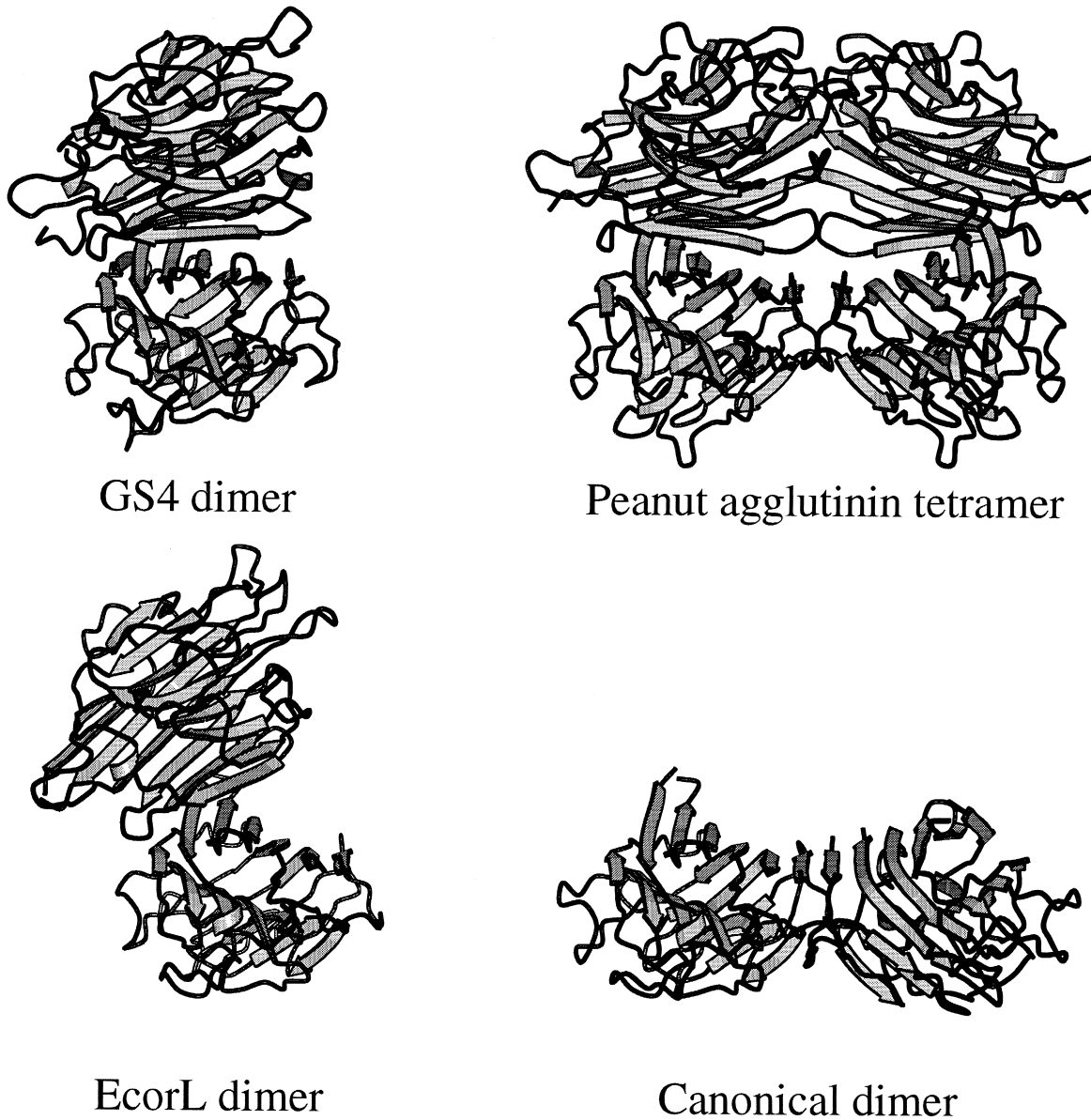


Fig. 4. The legume lectin dimers and the PNA tetramer. (a) the canonical dimer. (b) the GS-IV dimer. (c) The dimer formed by *Erythrina corallodendron*. (d) the peanut agglutinin tetramer. For the ease of comparison, one of the monomers is always shown in an identical orientation.

low pH is the canonical dimer, and at neutral or higher pH two such dimers associate with their back sheets [30,43]. This structure is shown in Fig. 5(a). Because of the concave course of the back β-sheets, the centre of the tetramer is formed by a large water-filled cavity of about $25 \times 25 \times 8$ Å^3. Several charged and ionisable side chains are buried upon tetramer formation, some of which are involved in salt-bridges. Interestingly, the tetramer interface also contains two pairs of arginines (Arg60 from each of the four subunits) in close contact. In the structure of demetallised concanavalin A at pH 5.0 they actually are involved a stacking interaction with each other [30].

All known crystal structures of concanavalin A have been obtained around pH 7 [20,44–47], except for demetallised concanavalin A and its monometallised Co^{2+}- and Zn^{2+}-derivatives, that were crystallised at pH 5 [30,31,48]. Although dimeric in solution, demetallised concanavalin A, through crystal packing interactions, forms in the crystal a tetramer very similar to that observed at pH 7.0. The number of van der Waals and hydrogen bonding contacts between the two dimers is, however, drastically reduced [48]. This reduction in inter-dimer interactions can largely be explained by the different protonation state of aspartates and glutamates at the interface.

The different crystal forms of concanavalin A (native, a mannose complex and two trimannoside complexes) and the almost identical lectin from *Canavalia brasiliensis* (two amino acid substitutions of which one on the interface between the two canonical dimers) have shown that the structure of the concanavalin A type tetramer is variable to some extent [45,49–51]. The two dimers may rotate several degrees relative to each other and may even show a small relative translation, leading to a different hydrogen bonding scheme on the dimer–dimer interface. In the concanavalin A–trimannoside complex [49], a small relative rotation of the two monomers in the canonical dimers is observed as well. This suggests the concanavalin A tetramer to be a dynamical structure in solution.

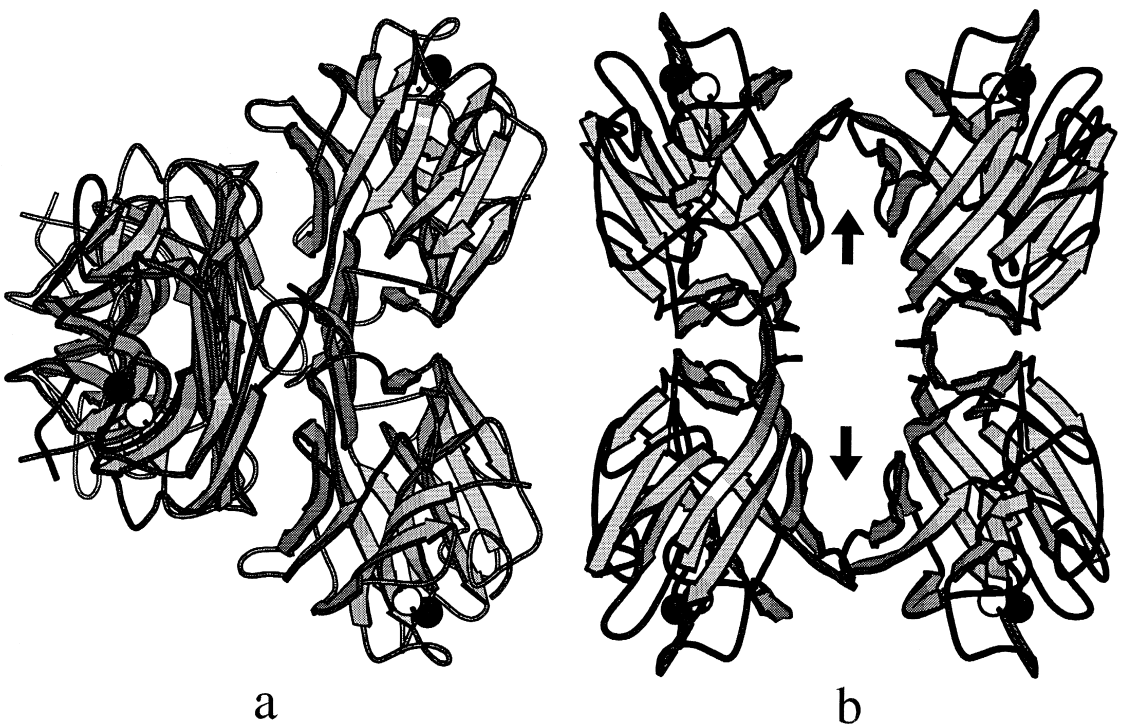


Fig. 5. Comparison of the tetramers formed by (a) concanavalin A and (b) PHA-L and soybean agglutinin. The presumed adenine binding site is indicated by arrows. Both tetramers consist of two canonical dimers.

3.4. The peanut agglutinin tetramer

Peanut lectin (*Arachis hypogaea* – PNA) was the second tetrameric lectin of which the structure was determined. Although its crystallisation was reported as early as 1982, a structure refined at 2.2 Å was only described recently [13,14]. Despite its homology with other known legume lectins, this structure could only be solved using isomorphous replacement. The peanut lectin tetramer can be considered to consist of two GS-IV type dimers (Fig. 4). Two monomers, one from each of these GS-IV type dimers, then associate in a fashion similar but not identical to the canonical dimer, resulting in the only known tetrameric protein that does neither contain 4-fold or 222 symmetry. However, no continuous 12-stranded β-sheet is formed along the dimer interface. Instead, the two N-terminal strands of the two monomers that associate in the canonical dimer-like fashion are intercalated by a series of 6 water bridges.

The PNA tetramer is thus unique among all homotetrameric proteins in the sense that its internal symmetry does not lead to a closed point group. It should also be noted here that peanut agglutinin is not a glycoprotein. Therefore, its unique quaternary structure cannot be due to steric effects of covalently bound carbohydrate and must be due to intrinsic properties of the protein itself as coded into its amino acid sequence. Modelling studies on the concanavalin A and peanut agglutinin tetramers [14] suggested that the deciding factor in the choice of the type of subunit association might be the amount of non-polar surface area buried upon oligomerisation.

3.5. The PHA/SBA tetramer: an adenine binding motif?

The tetramers of soybean agglutinin (SBA) [15] and the leuko-agglutinin from *Phaseolus vulgaris* (PHA-L) [8] have in common with concanavalin A that they also consist of two canonical dimers. The resulting tetramers are, however, distinctly different (Fig. 5(b)). Tetramer formation is not pH dependent for PHA-L and SBA. In the PHA-L and SBA tetramers, the two curved 12-stranded back sheets interact with their two outmost strands, creating a large channel in the middle of the tetramer. This channel probably contains the disordered C-terminus of these lectins and protects it from proteolyic degradation. The interface formed by these two outmost strands consists mainly of a number of relatively short side chains that intercalate in a zipper-like fashion, an architecture which is unusual for β-sheet packing.

Many legume lectins contain, in addition to their carbohydrate binding site, one or more binding sites for hydrophobic ligands [52–54]. Of special interest here is the binding of adenine and adenine-related plant hormones [55–59]. Not all legume lectins contain an adenine binding site, but those that do recognise adenine and certain adenine-related cytokinins with an affinity of 10^{-5}–10^{-6} M, i.e. several orders of magnitude stronger than their affinities for monosaccharides and even oligosaccharides [57–59]. Structural data on these lectin hormone interactions is limited, but it has been determined that the stoichiometry is either one or two binding sites per tetramer [54,58]. Affinity labelling experiments identified amino acid residues neighbouring the binding site and thus allowed for a crude localisation of the adenine binding site on the PHA-L tetramer as a 2-fold symmetric binding site on the interface between different monomers in the tetramer, inside the central hole of the tetramer [8]. The broad location of this binding site is shown in Fig. 5(b).

The PHA/SBA tetramer has recently been observed in the crystal structures of *Vicia villosa* lectin [147] as well as for the seed lectin of *Dolichos biflorus* and lectin II of *Ulex europaeus* (our unpublished results). Of these, the *Dolichos biflorus* lectin is known to bind adenine, while in the case of *Vicia villosa* and *Ulex europaeus* adenine binding has never been investigated. The PHA/SBA tetramer seems thus to be the default or canonical tetramer, while the concanavalin A and peanut agglutinin tetramers are the exceptions.

3.6. Post-translational processing

Legume lectins are generally synthesised as a precursor, which subsequently (after an initial removal of an N-terminal signal peptide) is post-translationally processed into a mature protein. This post-translational processing may consist of proteolytic cleavage of the precursor chain [60], C-terminal trimming [62–64], removal of covalent carbohydrate [65,66] and even ligation of the original C- and N-termini [67,68]. Depending on each specific lectin, the differ-

ent precursors may or may not possess carbohydrate binding activity.

Undoubtedly, the most complex and intriguing type of processing is the chain of events leading to a circularly permuted structure in concanavalin A [67,68]. After removal of a signal peptide followed by deglycosylation [69], a loop segment is removed in the middle of the sequence, presumably by an asparagine-specific protease [70,71]. In about two thirds of the molecules, the original N- and C-termini are then ligated, resulting in a circularly permuted protein [67,68]. This relegation was recently shown to be catalysed by jack bean asparaginyl endopeptidase [72].

Another well known form of post-translational processing is the one giving rise to the so-called two-chain lectins of the *Vicieae* tribe. Here, an amino acid segment of 6 residues is cut out of the pro-lectin, resulting in a protein containing a short C-terminal α-chain and a long N-terminal β-chain. Examples are the lectins from pea, lentil, broad bean and *Lathyrus* species [60,61].

Less well known is that many so-called single chain lectins that do not undergo the above processing contain ragged C-terminal ends, as determined by mass spectrometry. Examples are soybean agglutinin, PHA-E and the lectins from *Dolichos biflorus* [62–64]. This phenomenon has also been observed for recombinant lectins expressed in *E. Coli* [64]. Some of these lectins were shown to consist of an apparent equimolar mixture of long and short variants. In the case of the *Dolichos biflorus* seed lectin and the closely related stem and leaf lectin, it has been suggested that C-terminal truncation of half of the subunits of the respective tetramer and dimer is essential for carbohydrate binding activity and that only the untruncated subunits recognise carbohydrates [73]. The molecular basis of this phenomenon is not known, but may be related to the quaternary structure of these proteins. A more detailed explanation probably will have to await the determination of the crystal structure of these two lectins.

3.7. Relationship between oligosaccharide specificity and quaternary structure

Many plant lectins, covering a wide range of carbohydrate specificities, have been shown to precipitate glycoproteins and branched multivalent oligosaccharides [74]. This precipitation process often leads to the formation of homogeneous cross-linked lattices of a crystalline nature and in some cases leads to macroscopic crystals [75–77]. The earlier mentioned structure of soybean agglutinin in complex with the biantennary analogue of the blood group I carbohydrate antigen Galβ1–4-GlcNAcβ[Galβ1–4-GlcNAcβ]-Galβ-O$(CH_2)_5COOCH_3$ is such an example [15]. This type of lectin–carbohydrate interaction is extremely specific as in mixed precipitation systems, distinct and homogeneous lattices can be observed, even if the individual monovalent constituents of the carbohydrate or glycoprotein are bound only with weak affinity [78]. This type of interaction thus represents a new source of binding specificity with a high biological relevancy [79,80]. Binding of multivalent lectins to the surface of cells often leads to cross-linking and aggregation of specific glycoprotein and glycolipid receptors, which in turn is associated with a variety of biological responses such as mitogenic activities and a variety of biological signal transduction processes.

The formation of homogeneous lattices can in part be explained by the multimeric nature of the lectins. As discussed previously, different lectins possess different types of quaternary structures, independent of their specificity for monosaccharides. The different spacings between the carbohydrate binding sites of the different lectins combined with specific distances between the different epitopes on multivalent ligands are probably the reason for this type of specificity. This is in agreement with observations on other interacting systems such as the interaction of cholera toxin B with the GM1 ganglioside in a lipid layer and actin interacting with actin-cross-linking proteins [79,80].

4. Molecular basis of carbohydrate specificity

In the following paragraphs, we will give a detailed description of the structural basis of mono- and oligosaccharide binding to legume lectins. Important similarities and differences between the different lectins will be highlighted. Because of the circular homology between concanavalin A and the other lectins and because the carbohydrate binding residues

are distributed over several loops, the equivalent residues for the lectins that are discussed below are given in Table 4. These equivalences are based on superpositons of the different coordinate files (obtained from X-ray crystallography or molecular modelling), except for Lima bean lectin, GS-II and *Ulex europaeus* lectin II, where no three-dimensional coordinates were available. For these three lectins, the equivalent residues were determined by aligning the published amino acid sequences.

4.1. Glc / Man specific lectins

4.1.1. Concanavalin A: an extended binding site

Concanavalin A, the lectin from Jack bean (*Canavalia ensiformis*) was the first legume lectin to be isolated [81], sequenced [82,83] and its three dimensional structure determined by X-ray crystallography [6,7]. The structure of a concanavalin A – methyl α-D-mannopyranoside complex at 2.9 Å resolution in 1989 provided the first detailed view on how a legume lectin specifically interacts with a monosaccharide [84]. This complex has recently been refined at 2.0 Å [46] and the interactions in the binding site are typical for all legume lectin carbohydrate complexes elucidated by X-ray crystallography to date. The essential interactions involve key hydrogen bonds with three extremely well conserved residues: an aspartate (Asp208) that is preceded by a *cis*-peptide bond and is held in place via a water bridge with a structural calcium ion, an asparagine (Asn14) that directly interacts with this calcium and the backbone NH of Arg228, which in most other legume lectin sequences is a glycine and which also interacts with the calcium ion via a water bridge with its backbone oxygen (Figs. 6 and 7). Also essential are van der Waals interactions between aromatic residues (Tyr12 in concanavalin A, but often also Phe and sometimes Leu or Cys) and the sugar ring. Further interactions involve a loop segment (Thr97–Glu102), of which the conformation and length is variable among the known legume lectin structures and that is not in contact with the calcium ion.

Conformational changes in the binding site of concanavalin A upon saccharide binding are small, in contrast to the large differences between holo- and apo-concanavalin A. Two differences between the complexed and the uncomplexed structures are nevertheless significant. The first one concerns a movement of the side chain of Arg228. In all structures of uncomplexed concanavalin A, this side chain partially blocks the monosaccharide binding site. This has been related to the apparent co-operativity of the binding of Manα1–3[Manα1–6]Man-α-Me by concanavalin A, which may have a dynamic origin, as is suggested by recent calorimetric studies [85–87]. It can however not be excluded that the observed blocking of the binding site by Arg228 results from crystal packing forces, since the crystal form of uncomplexed concanavalin A is densely packed. Most other legume lectins have a glycine at the position of Arg228. The second conformational change concerns a small movement (about 0.5 Å) of the loop Thr97–Glu102. Strikingly, in all other legume lectins where both a complexed and uncomplexed structure are available, small movements (< 1.0 Å) of the corresponding loop are observed upon carbohydrate binding.

The effect of the presence of a bulky hydrophobic substituent on O1 of mannose and glucose has been investigated [88,89]. In the complexes of concanavalin A with 4′-nitrophenyl α-D-mannopyranoside and 4′-nitrophenyl α-D-glucopyranoside the nitrophenyl group hydrogen bonds to the side chain hydroxyl of Tyr100, and makes favourable hydrophobic interactions with side chain atoms of Tyr12, Leu99 and Tyr100 (Fig. 7).

Concanavalin A, as well as the closely related lectin from *Dioclea grandiflora*, has its highest affinity for the trimannosyl core Manα1–3[Manα1–6] Man found in *N*-glycans [90,91]. Recently, two groups independently determined the structure of a complex between concanavalin A and the trimannoside [49,92]. The terminal mannose on the α1–6 arm is bound in the monosaccharide binding site, in a way similar as observed in the concanavalin A – methyl α-D-mannopyranoside complex. The reducing core mannose and the α1–3 terminal mannose are bound in an extension of this primary site (Fig. 8(a)). The core mannose interacts with the side chains of Tyr100 and Asp16 via its O4 and O2 oxygens respectively. These results are consistent with recent thermodynamic data on the concanavalin A – trimannosyl system [93,94].

Both published structures agree with each other on the interactions between the core mannose and the

Table 4
Residues involved in carbohydrate recognition

Concanavalin A	*Vicieae* lectins	*Erythrina corallo*-dendron lectin	*Soybean* agglutinin	*Dolichos biflorus* seed lectin	Lima bean lectin	Peanut agglutinin	GS-IV	Function
Tyr12	Phe123β	Phe131	Phe128	Leu127	Cys127	Tyr125	Trp133	a,b
Pro13	Tyr124β	Ser132	Arg129	Ser128	His128	Ser126	Ile134	b
Asn14	Asn125β	Asn133	Asn130	Asn129	Asn129	Asn127	Asn135	a
Thr15	Ala126β	–	–	Ser130	Leu130	Ser128	Lys136	b
Asp16	Ala127β	–	–	Gly131	Asp131	Glu129	Asp137	b
Thr15–Ile17	Ala126β–Trp128β	Pro134–Trp135	Ser131–Trp132	Ser130–Trp132	Leu130–Asp131	Ser128–Asn131	Lys136–Trp138	c
Ile17	Trp128β	Trp135	Trp132	Trp132	Trp132	Tyr130	Trp138	c
Thr97–Glu102	Thr28α-Ala33α	Thr216-Glu224	Thr212–Glu219	Thr212–Glu220	Ser212–Glu220	Ser211–Gln216	Val221–Glu225	e
Leu99	Ala30α	Ala218	Leu214	Leu214	Leu214	Gly213	Tyr223	d
Tyr100	Glu31α	Gln219	Asp215	Ser215	Asn215	Gly214	Asp224	b,d
Asp208	Asp81β	Asp89	Asp88	Asp85	Asp85	Asp83	Asp89	a
Thr226	Gly97β	Gly105	His104	Asn101	Asn101	Gly102	Tyr105	c
Gly227	Gly98β	Tyr106	Ala105	Gly102	Gly102	Gly103	Gly106	c
Arg228	Gly99β	Gly107	Gly106	Gly103	Gly103	Gly104	Gly107	a
Leu229	Tyr100β	Tyr108	Tyr107	Tyr104	Phe104	Thr105	Phe108	b,d

[a] Residues essential for monosaccharide binding: the conserved Asp–Asn–Gly(Arg) triad and an hydrophobic residue in van der Waals contact with the sugar ring.
[b] Residues systematically implicated in the higher affnity for oligosaccharides (such as the trimmanose core in the case of concanavalin A).
[c] Residues identified by NMR, modelling and crystallography to be implicated in interactions with an aglycon (*Vicieae* lectins), an *N*-acetyl group (*Dolichos biflorus*) or a fucose residue (*Erythrina corallodendron*). In the case of *E. corallodendron*, Pro134–Trp135 is also involved in the recognition of a dansyl group.
[d] Hydrophobic pocket that in the case of ConA interacts with aglycons on O1.

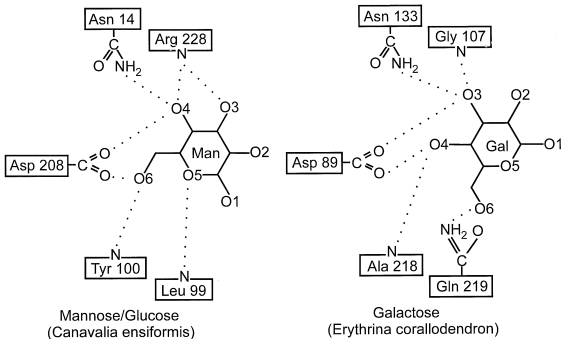


Fig. 6. Monosaccharide recognition. (left) Schematic diagram of the binding of mannose to concanavalin A. (right) Similar diagram for the recognition of galactose by the lectin from *Erythrina corallodendron.*

protein, but differences are seen for the mannose on the α(1–3) arm. In the structure reported by Naismith and Field, all trimannose molecules are bound in the same conformation, while Loris and co-workers observed two conformations for the α(1–3) connection. Interestingly, these two conformations resulted from the same hydroxyls interacting with the protein. The existence of two conformations in a ligand that is considered a specific binder is unusual and was therefore extensively verified using the experimental data. Despite the convincing electron density for this situation, no satisfying explanation could be found as to why this is the case. In fact, the trimannose molecule with the alternative conformation is not involved in crystal packing interactions and seems not to be prevented in any way to adopt the same conformation as the other trimannose molecules in the asymmetric unit.

Concanavalin A also has a high affinity for structures containing Manα1–2-Man linkages. No crystallographic or NMR data is available, but molecular modelling has been attempted with the disaccharide Manα1–2-Man [95]. From this study, it was concluded that the non-reducing mannose most likely occupies the monosaccharide binding site, while the reducing mannose interacts with the side chains of Asn14 and Arg228.

4.1.2. Mimicri between carbohydrate and peptide recognition.

A couple of years ago, it was reported by two groups that peptides binding to the carbohydrate binding site of concanavalin A could be isolated from a random peptide library [148,149]. These peptides bind to concanavalin A with an affinity similar to that of methyl α-D-glucopyrannoside and inhibit dextran precipitation of concanavalin A. Conversely, monosaccharides such as α-D-mannopyrannoside inhibit peptide binding. These peptides, that contain the consensus sequence Tyr–Pro–Tyr, do not bind to other closely related Glc/Man specific legume lectins indicating interactions both with the monosaccharide binding site and some surrounding residues that are not conserved within the Glc/Man specificity group. On the other hand, anti-DVFYPYPYASGS antibodies recognise carbohydrate ligands and anti-α-D-man-

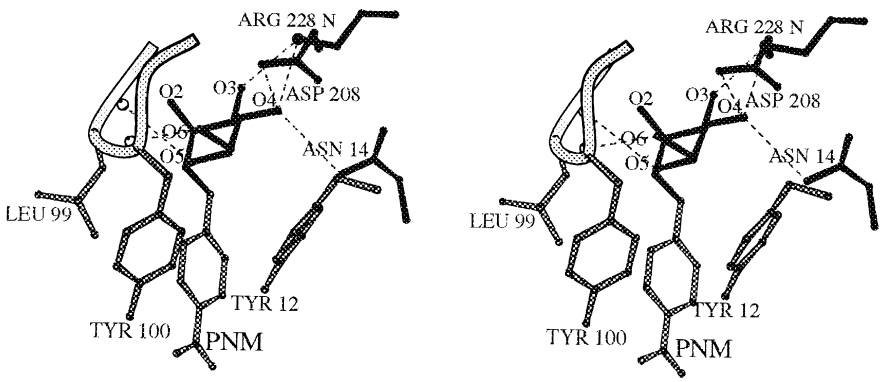


Fig. 7. Stereoview of the hydrophobic binding site of concanavalin A adjacent to the monosaccharide binding site. The saccharide moiety hydrogen bonds to the side chains of Asn14 and Asp208, as well as to the main chain NH groups of Arg228, Leu99 and Tyr100. Further important van der Waals interactions are observed between the sugar ring and the side chain of Tyr12. The aglycon of 4′-nitrophenyl α-D-mannopyranoside (labeled PNM) is bound in a hydrophobic cavity formed by Tyr12, Leu99 and Tyr100.

nopyrannoside antibodies recognise the peptide DVFYPYPYASGS that also binds to the carbohydrate binding site of concanavalin A [150]. This situation is analogous to the mimicri found between the cytokeratin peptide SFGSGFGGY and *N*-acetyl β-D-glucosamine [151]. Also here, antibodies against GlcNac and the peptide are interchangeable and the peptide is recognised by some GlcNac-specific lectins such as wheat germ agglutinin.

4.1.3. Vicieae lectins: variations on a theme

Crystal structures of the lectins from fava bean (*Vicia faba*), pea (*Pisum sativum*), lentil (*Lens culinaris*) and two isolectins from *Lathyrus* (*Lathyrus ochrus* – LOL I and LOL II) with monosaccharides have shown that these very homologous lectins (over 85% pairwise sequence identity) bind mannose and glucose in a way essentially identical to concanavalin A [96–99].

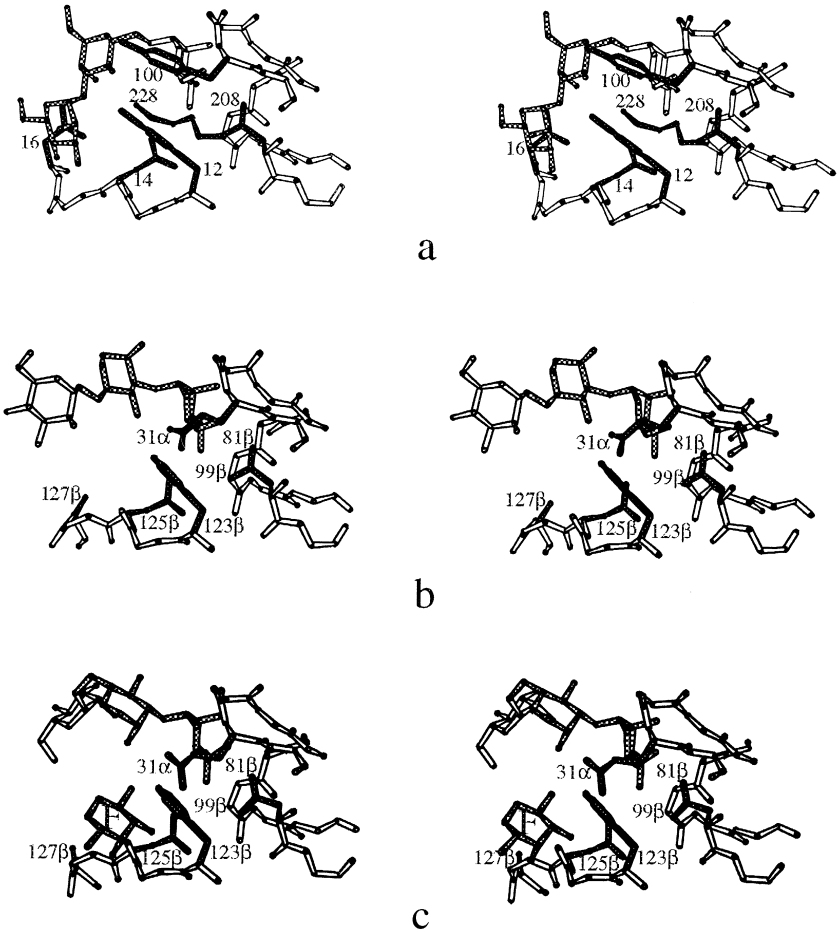


Fig. 8. Oligosaccharide specificity in Glc/Man specific lectins. (a) Stereo figure of the trimannose core bound in the binding site of concanavalin A. The 1–6 terminal mannose that occupies the monosaccharide binding site is shown in white, the other two mannose residues in grey. The amino acid side chains that hydrogen bond to either of the three mannose residues (Tyr12, Asn14, Asp16, Tyr100, Asp208 and Arg228) are shown in black. (b) Similar view of the trimannose part of an octasaccharide bound to *Lathyrus ochrus* lectin. Colour-coding is identical to that used in (a). The 1–6 teminal mannose and the core mannose stick into the solvent, while the 1–3 terminal mannose occupies the monosaccharide binding site. The black residues are the equivalents of those of concanavalin A: Phe123β, Asn125β, Ala127β, Glu31α, Asp81β and Gly99β). (c) Trimannose part of a glycopeptide bound to *Lathyrus ochrus* lectin. The situation is identical to (b), except for the different conformation of the trisaccharide. In addition, the specificity-determining fucose is also shown in grey and labeled ''F''. Its position roughly coincides with that of the 1–3 terminal mannose in (a) and it interacts with the equivalent amino acid side chains.

In the case of LOL I, the effect of an aglycon on enhanced affinity was studied in the complexes with *N*-acetylmuramic acid and a *N*-acetylmuramyl dipeptide (Mur-D-Ala-D-iGln – MDP). It was found that the aglycon of both compounds interacts with the side chains of Tyr100β and Trp128β and with residues Gly97β–Gly98β of LOL [100]. These are the same residues that according to a recent molecular modelling study are responsible for the enhanced affinity of lentil lectin for 2-*O*- and 3-*O*-*m*-nitro-benzyl derivatives of mannose and glucose as illustrated in Fig. 9(a) [101]. This hydrophobic subsite is different from the one described above for concanavalin A, which cannot bind mannoses or glucoses containing substituents on O3 [102].

The binding of sugar residues extending from the mannose or glucose in the monosaccharide binding site is however fundamentally different from what is seen in concanavalin A. Sugar residues immediately attached to the mannose or glucose in the monosaccharide binding site are either not visible as in the complex of pea lectin with the trimannoside core [98], or only interact with the protein via an extended water network, as found in a sucrose complex of lentil lectin [99] and the complex of *Lathyrus* lectin with a trisaccharide [103]. This is in agreement with the affinities of lentil, pea and *Lathyrus* lectin for these oligosaccharides being similar to their affinities for α-D-mannopyranoside. In all cases, the carbohydrates are in a low (but not necessary the lowest) energy conformation [105] and are involved in or located close to crystal lattice interactions. Crystal packing forces nevertheless seem to have little influence on the interaction between the lectin and the oligosaccharide, as shown in an extensive study of the lentil lectin–sucrose interaction that compared results from crystallography, solution NMR and molecular modelling [99]. The conformation of the sucrose molecule, which is known to be highly flexible in solution, as well as the mode of interaction were found to be essentially identical when studied with the three techniques.

In contrast to concanavalin A, the *Vicieae* lectins have not significantly enhanced affinity for the trimannose core over α-D-mannopyranoside [91,106]. This can be explained by comparing the trimannoside complex of concanavalin A with the different oligosaccharide complexes of *Lathyrus* lectin [103,104]. LOL (and also pea and lentil lectin) lacks the functional groups that in the concanavalin A trimannoside structures interact with the core mannose: Tyr12 and Tyr100 in concanavalin A are substituted by Phe123β and Glu31α in LOL. As a consequence, the trimannoside cannot bind in the same orientation in LOL as in concanavalin A [92]. This is illustrated in Fig. 8(b) and (c). Instead, the α1–3 linked mannose binds in the monosaccharide binding site. The conformations of the α1–3 and α1–6 linkages in the LOL complexes are variable and depend on the exact nature of the longer oligosaccharide [100].

Lectins from the *Vicieae* tribe show a high affinity for *N*-acetyllactosamine-type glycans that have a fucose residue α1–6 linked to the N-linked *N*-acetylglucosamine [91,106]. The structural basis of this specificity is explained by two recent crystal structures of LOL complexed with fragment N2 of human lactotransferrin and with an isolated biantennary glycopeptide [107]. The fucose residue was found to be bound in a subsite, interacting with Phe123β and Glu31α, i.e. the same residues that in concanavalin A are responsible for its specificity for the trimannose core. Indeed, as shown in Fig. 8(c), the position of this fucose roughly coincides with that of the terminal 1–3 mannose residue in the concanavalin A–trimannose complexes. In absence of the fucose, the affinity of the lectins from the *Vicieae* tribe (LOL, favin, lentil, lectin and pea lectin) for the lactotransferrin-derived oligosaccharide is somewhat lower, but still about 100-fold higher than for the monosaccharides [91]. This is explained by the crystal structure of a LOL–octasaccharide complex [104], in which the non-fucosylated lactotransferrin-derived octasaccharide adopts a conformation upon binding to LOL that is different from the one adopted in complex of LOL with the fucosylated biantennary glycopeptide [107]. Recently, it was shown that both the fucosylated and the non-fucosylated biantennary glycans are highly flexible molecules and that upon complex formation, the *Vicieae* lectins select out a conformation that is already present in solution [155]. The differences in conformation of the fucosylated and non-fucosylated biantennary oligosaccharides bound to LOL as seen in the different crystal structures are therefore not due to inherently different solution conformations of these oligosaccharides, but rather due to a different lectin-mediated selection of a

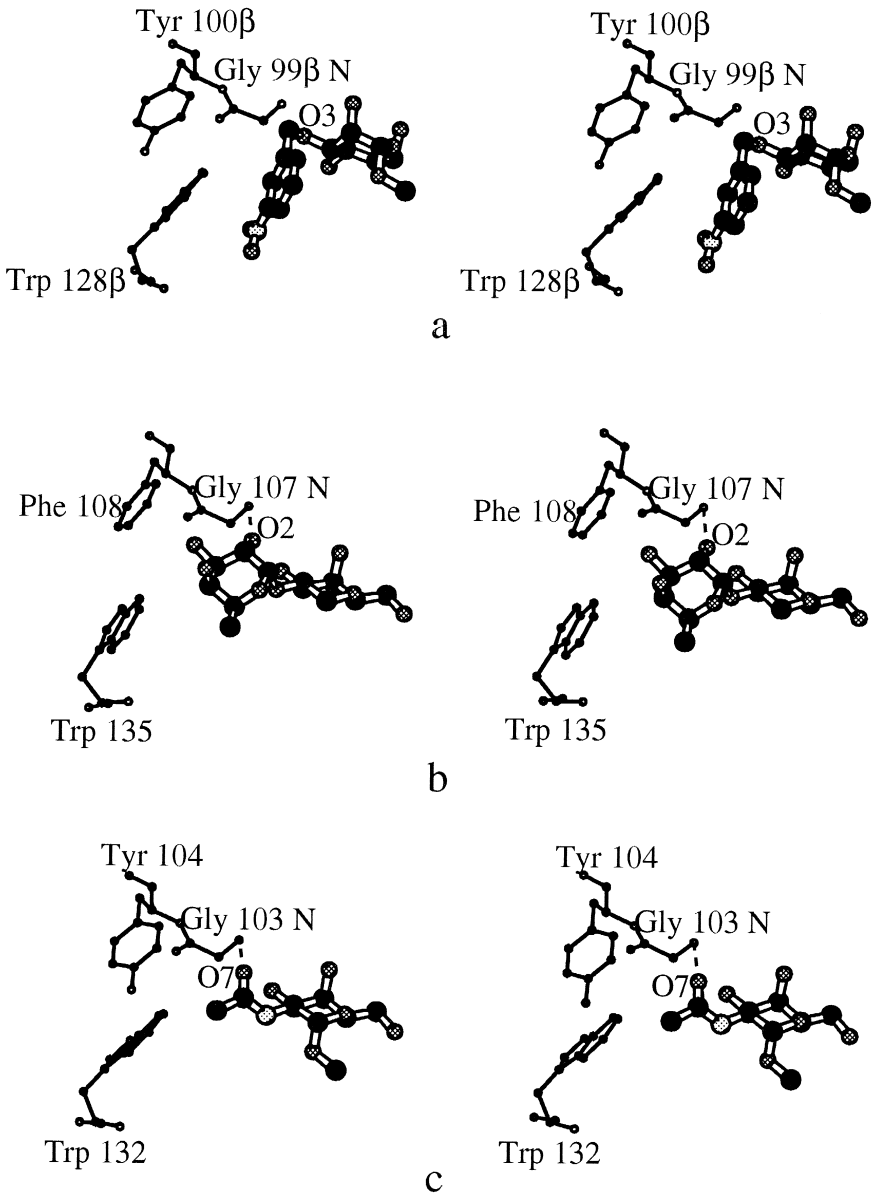


Fig. 9. A multi-purpose subsite. The same residues (a tryptophane, a tyrosine and a glycine) are responsible for high affinity binding of substituted monosacchrides. (a) The nitrophenyl group of 3-*O*-nitrophenylmannose in complex with lentil lectin. (b) The fucose moiety of Fucα1–2-Galβ1–4-GlcNAc complexed to the *Erythrina corallodendron* lectin (The GlcNAc residue is not shown for clarity). (c) The *N*-acetyl group of GalNAc complexed with the *Dolichos biflorus* seed lectin. A similar subsite was found four these three substituents using modelling calculations, and in the case of *Dolichos biflorus* also by NMR. The tyrosine and tryptophane are conserved in many, but not all legume lectins. Notable exceptions are concanavalin A and peanut agglutinin.

particular conformation depending on the presence or absence of fucose.

4.1.4. Other Man / Glc specific lectins

The above two fine specificity groups within the Man/Glc specific legume lectins are the only ones for which crystallographic information is available. There are, however, other Man/Glc specific lectins that exhibit fine specificities different from those already described, notably lectins from sainfoin (*Onobrychis vicifolia*), the Japanese pagoda tree (*Sophora japonica*) and from the West African

legume *Bowringia milbraedii* [108,109]. Of these, only for the *B. milbraedii* lectin has a detailed study on sugar specificity been published [110]. This lectin differs from the other Man/Glc specific lectins in that it is only poorly inhibited by glucose, and not at all by methyl β-D-glucopyranoside and has a high affinity for structures containing Manα1–2-Man linkages. *B. milbraedii* on the other hand does not seem to have a specific binding site for the trimannoside core Manα1–3[Manα1–6]Man [110].

4.2. Gal / Galnac specific lectins

4.2.1. Structural features of galactose complexes

The amount of structural data available for the Gal/GalNAc specific legume lectins is also growing steadily. At present, the structures of four legume lectins with a galactose residue in the monosaccharide binding site have been refined: *E. corallodendron* lectin in complex with lactose [16], peanut agglutinin in complex with lactose [14] and the T-antigen disaccharide [152], *Griffonia simplicifolia* lectin IV (GS-IV) complexed with the Lewis b tetrasaccharide Fucα1–2-Galβ1–3[Fucα1–4]GlcNAcβ-Me [17] and a cross-linked lattice between soybean agglutinin (*Glycine max*) and the biantennary pentasaccharide analogue of the blood group I carbohydrate antigen Galβ1–4-GlcNAcβ[Galβ1–4-GlcNAcβ]-Galβ-$O(CH_2)_5COOCH_3$ [15]. Of these complexes, the GS-IV complex is special in the sense that GS-IV does not bind free galactose or GalNAc [111]. GS-IV thus belongs sensu stricto to the complex specificity group and will be discussed in detail later. From the structures of *E. corallodendron* lectin, peanut agglutinin and soybean agglutinin, it can be seen that the hydrogen bonding network in the binding site in the presence of galactose is similar to the one found for the mannose-binding lectins. Since mannose and galactose differ in the orientation of their C4 hydroxyl (axial or equatorial), the orientation of galactose in the binding site needs to be different from the one of mannose, as first pointed out by Shaanan et al. [16] and further discussed by Sharon [153]. As can be seen in Fig. 6, this results in the conserved Asp–Asn–Gly triad interacting with O4 and O6 in the Glc/Man specific lectins and with O3 and O4 in the Gal specific lectins. A similar situation is found for glucose bound to the glucose/galactose binding protein from *E. coli* [112], as compared to the galactose orientation in the L-arabinose binding protein [113].

The mode of binding of galactose is essentially identical in the structures of *E. corallodendron* lectin and soybean agglutinin. In neither of these structures, direct hydrogen bonds are observed with another carbohydrate residue than the galactose in the monosaccharide bonding site. A major difference between *E. corallodendron* lectin and soybean agglutinin on the one hand and peanut agglutinin on the other involves the loop Thr216–Glu224 (*Erythrina* numbering), the conformation of which is strikingly different in the peanut structures [14]. Although similar hydrogen bonds are formed with the conserved Asp–Asn–Gly triad in peanut agglutinin, the exact position of the galactose residue in the monosaccharide binding site of peanut agglutinin is shifted somewhat on the peanut agglutinin structure compared to soybean agglutinin and *E. corallodendron* lectin.

All available crystal structures contain a galactose in their monosaccharide binding site and therefore do not explain the higher affinity of most of these lectins for GalNAc. In order to understand this, Imberty and co-workers modelled the (strictly) GalNAc specific seed lectin from *Dolichos biflorus* (DBL) and used this structure to dock various saccharides, including GalNAc [114]. It was proposed that the carbonyl oxygen of the *N*-acetyl group of GalNAc hydrogen bonds to the main chain NH group of Gly103, while the methyl group forms favourable van der Waals interactions with the aromatic side chains of Tyr104 and Trp132 (Fig. 9(b)). These results have been corroborated by a recent NMR study on the interaction between DBL and the Forssman pentasaccharide GalNAcα1–3-GalNAcβ1–3-Galα1–4-Galβ1–4-Glc [115]. In this study, intermolecular NOE's between the protein and the carbohydrate ligand could be observed for the terminal GalNAc residue that indeed suggest that the *N*-acetylamine group interacts with Tyr104 and Trp132.

The blood group A trisaccharide GalNAcα1–3[Fucα1–2]-Gal occurs in solution in two main conformational populations. In the same modelling study involving the *Dolichos biflorus* seed lectin [114], it could not be determined with certainty which of the two conformations binds to the lectin. This ambiguity was resolved using an NMR study on the DBL-blood group A trisaccharide complex [116]. Combination of

the modelling and NMR studies indicate that when bound to DBL, the O4 hydroxyl of the fucose residue hydrogen bonds to the main chain carbonyl of Ser128. This residue corresponds to Tyr124β in the *Vicieae* lectins and Pro13 in concanavalin A. It is therefore of relevance that in the concanavalin A–trimannoside complexes, the carbonyl of Pro13 also hydrogen bonds with the 1–3 terminal mannose residue. Further van der Waals contacts are observed between the fucose and Leu127, the equivalent of Phe123β in the *Vicieae* lectins and Tyr12 in concanavalin A. These results strongly suggest similar interactions not only in the monosaccharide binding sites of lectins with different specificities, but also in a conserved subsite responsible for high affinity binding of oligosaccharides.

The structural basis of oligosaccharide specificity was also investigated for the *E. corallodendron* lectin [117]. This lectin has a high affinity for Fucα1–2-Galβ1–4-GlcNAcβ-terminated glycosphingolipids. Modelling studies with this trisaccharide suggest that the α1–2-linked fucose binds in a cavity extending from the monosaccharide binding site. In this cavity, the fucose residue forms hydrophobic contacts with Pro134 and Trp135, as well as hydrogen bonds with the NH2 group of Asn133 (to the 2-OH of fucose as well as to the glycosidic linkage), NH of Gly107 (to the 2-OH of fucose) and the hydroxyl group of Tyr108 (to the 3-OH of fucose). These residues correspond to those that form the so-called ''hydrophobic binding site'' adjacent to the monosaccharide binding site in both the Man/Glc (see above) and Gal (see below) specific lectins (Fig. 9(c)).

Very recently, the complex of peanut agglutinin with the T-antigen disaccharide (Galβ1–3GalNac) provided the first crystallographic data of a galactose specific legume lectin recognising an oligosaccharide that binds with an affinity that is significantly better than the lectin's affinity for Gal or GalNAc [152]. This structure is remarkable because it suggests that the 20-fold increase in affinity compared to lactose is entirely due to water-mediated protein–carbohydrate interactions, the amount of direct protein–carbohydrate hydrogen bonds and non-polar contacts being identical (Fig. 10). While specific water-mediated recognition has been observed previously in protein–DNA interactions, the peanut agglutinin-T-antigen disaccharide complex for the first time shows that water-mediated specificity is also possible for carbohydrate recognition.

4.2.2. Site directed mutagenesis

The galactose specific lectins are also the only ones for which their specificity has been successfully modulated by site directed mutagenesis. The lectin from *E. corallodendron* has a 500-fold higher affinity for galactoses containing a bulky substituent on C2, such as dansylamido (NDns) compared to the unsubstituted monosaccharide [118], and also a 2-fold higher affinity for GalNAc than for galactose [119]. In contrast, the affinity of peanut agglutinin for GalNAc is 5 times lower than for galactose, and its affinity for MeβGalNDns is only 20-fold higher than

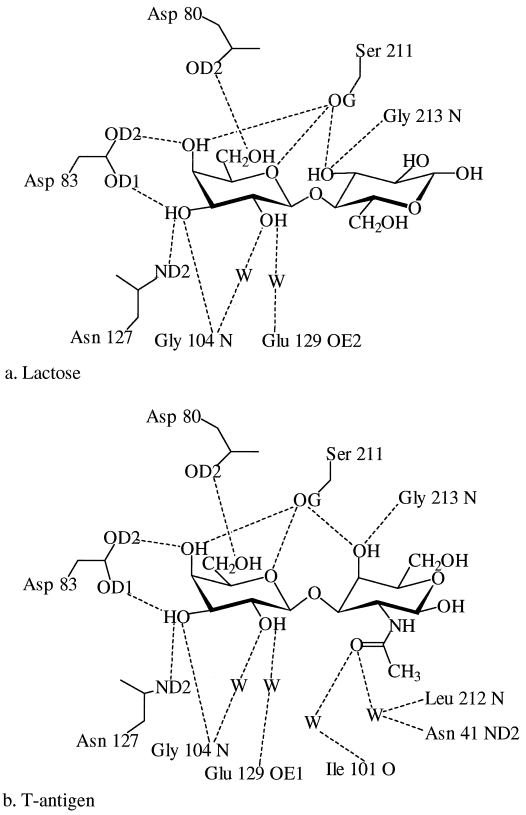


Fig. 10. Schematic diagrams of the binding of lactose (a) and the T-antigen disaccharide (b) to peanut agglutinin. The main difference between both complexes are the additional water-mediated interactions between the protein and the *N*-acetyl group of the GalNAc residue in the T-antigen disaccharide. Hydrogen bonds are shown as dashed lines and water molecules are indicated by ''W''.

for galactose [120]. The structure of the *E. corallodendron* lectin revealed an hydrophobic cavity near the C2 carbon of the bound galactose which was thought to be responsible for its affinity for galactoses containing substituents on C2. This cavity is formed by Tyr108, Pro134 and Trp135. Arango and co-workers substituted the dipeptide Pro–Trp by the corresponding tetrapeptide Ser–Glu–Tyr–Asn found in the related sequence of peanut agglutinin [121]. This resulted in a recombinant lectin with an unchanged affinity for galactose, a slightly diminished affinity for GalNAc and a markedly lower affinity for MeβGalNDns. This change in specificity could be rationalised as resulting from steric hindrance by the two additional amino acids present in the cavity region of the mutant.

Similar results were obtained by Jordan and Goldstein, who modified the binding site of lima bean (*Phaseolus lunatus*) lectin, another Gal/GalNAc specific lectin for which at the moment no crystal structure is available [122]. This lectin has its highest affinity for the blood group A immunodominant trisaccharide GalNAcα1–3[Fucα1–2]Galβ1-R. Two lima bean lectin mutants showed altered specificity for C2 substituents on the galactose residue of this trisaccharide: the mutant His128Pro recognised GalNAcα1–3-GalNAc while the Cys127Tyr mutant showed specificity for both GalNAcα1–3-Gal and GalNAcα1–4-Gal (with a necessity for a free hydroxyl group at the C2 position of Gal). It is of interest to note here that Cys127 corresponds to Tyr12 in concanavalin A and Phe123β in *Lathyrus ochrus* lectin, suggesting again that the same residues or loops that are responsible for determining the oligosaccharide specificity in the Man/Glc specific lectins have a similar role in the Gal/GalNAc specificity group. Interestingly, the same is true for the three residues that make up the above mentioned hydrophobic cavity. Both Tyr108 and Trp135 are conserved in the lentil, pea and *Lathyrus* sequences (as Tyr100β and Trp128β) and were identified as the amino acids responsible for the higher affinity of mannoses and glucoses with hydrophobic substituents on C2 or C3 [101]. In the *Dolichos biflorus* seed lectin, these residues correspond to Tyr104 and Trp132, which were proposed to be responsible for interacting favourably with the *N*-acetyl group of GalNAc [114].

4.2.3. Determinants of galactose and mannose specificity

As has been discussed, both the Glc/Man and the Gal/GalNAc specificity group lectins seem to make similar hydrogen bonds between the monosaccharide and the protein upon binding. Indeed, the key hydrogen bonding partners belong to the conserved core of residues identified by Young and Oomen [123]. Therefore, despite the well-known directionality of hydrogen bonds, hydrogen bonding does not seem to be an important determinant for the selectivity of the recognition process.

When the structures of the Glc/Man specific complexes are optimally superimposed onto the complex of the *E. corallodendron* lectin with lactose, it becomes clear that also the shapes of the monosaccharide binding sites are well conserved. The only exception seems to be the loop from Thr97 to Glu102 in concanavalin A (corresponding to Thr216–Glu224 in *Erythrina* and Thr28α–Ala33α in the lectins from *Lathyrus*, lentil and pea). Clear steric clashes with these loops are observed if a monosaccharide that is normally not recognised is placed into the binding site (Fig. 11). Since these loops mainly hydrogen bond to the bound carbohydrate with their backbone NH and CO groups, one may assume that the main determinant for choosing between mannose or galactose specificity is the length and conformation of this loop, but not directly by the individual amino acid side chains (in the case of *E. corralodentron* lectin, hydrogen bonds between side chains of this loop and lactose are observed, but proven to be dispensable by site directed mutagenesis – see below). We will therefore call this loop the ‘‘monosaccharide specificity loop’’, as we will show in the following sections that it also plays a similar role in determining fucose, chitobiose and complex specificity. Essentially identical conclusions were also drawn in a recent comparative modelling study [124]. These authors found that the length of the monosaccharide specificity loop (termed ‘loop D’ in Ref. [124]) is invariant in the Glc/Man specificity group and that there is a correlation between the length of this loop and monosaccharide specificity.

This idea is further corroborated by recent mutagenesis studies on the *E. corallodendron* lectin [125], the GlcNAc specific lectin II from *Griffonia simplicifolia* (GS-II) [126], a bark lectin from *Robinia pseu-*

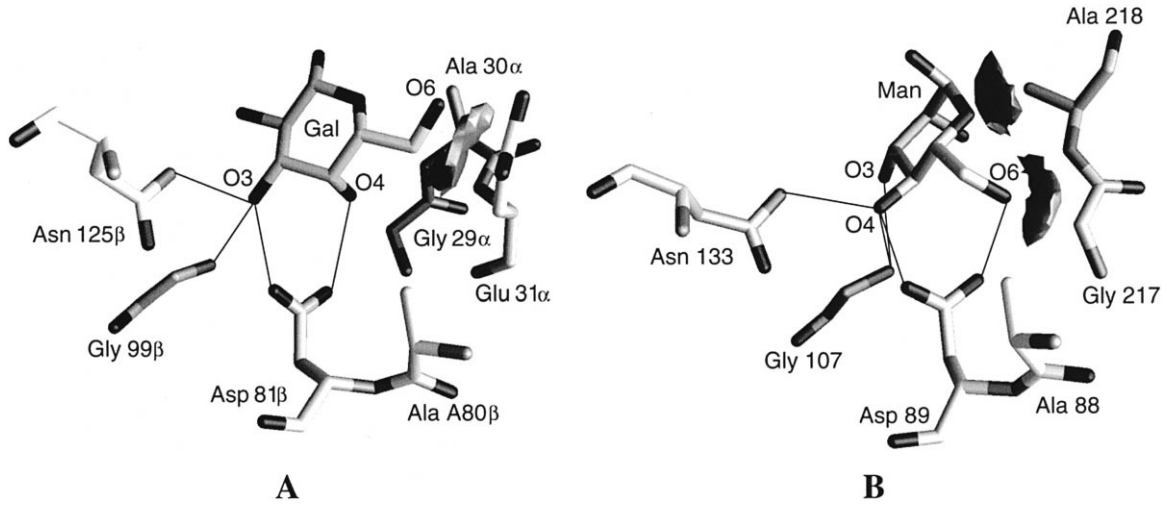


Fig. 11. Comparison of the monosaccharide binding sites of the lectins from *Lathyrus ochrus* and *Erythrina corallodendron*. (a) The combining site of LOL is show together with a docked galactose. The galactose was positioned by calculating the best least-squares superposition of the invariant monosacchride binding residues (Asp89, Gly107, Phe131 and Asn133) residues involved in metal binding from the *Erythrina corallodendron* lectin on the corresponding residues of LOL. This transformation was then applied to the galactose residue present in the co-ordinate set from *Erythrina corallodendron*. The resulting clashes between the saccharide and the lectin were visualised as a semi-transparent surface using the program GRASP [157] and are entirely confineded to the loop from Thr28α to Glu31α. (b) similar view of mannose docked into the binding site of the *Erythrina corallodendron* lectin. van der Waals clashes are observed between the docked mannose and the loop from Thr216 to Glu224.

doacacia [127] and pea lectin [154]. In these studies, muting the conserved Asp and Asn invariantly results in a complete loss of carbohydrate binding activity. Also, in the *E. corallodendron* lectin, the activity of mutants involving Phe131 (which stacks upon the galactose ring in the monosaccharide binding site) depended upon the nature of the replacement residue: replacement with the isosteric and hydrophobic Tyr is allowed, but the smaller Val and Ala lead to an inactive protein [125]. Similar results were also found for substitutions of Tyr124 in GS-II [126]. Point mutations in the monosaccharide specificity loop on the other hand do not lead to inactivation in the case of *E. corallodendron* lectin [125–127,154], but deletion mutants in this loop resulted in inactivation of GS-II.

The similarities between the Man/Glc and Gal specificity groups extend also towards oligosaccharide recognition. As described in the previous paragraphs, two sets of residues appear to form recurrent themes. The first set are those residues that form the so-called hydrophobic binding site, which is present in members of both specificity groups. This subsite is not only responsible for the enhanced affinity for monosaccharides substituted with hydrophobic aglycons, but may also play a role in binding the *N*-acetylamine group of GalNAc and even in oligosaccharide specificity in the case of *E. corallodendron*, where they have been suggested to form a fucose subsite. The second set involves Tyr100 and Asp16 in concanavalin A and their equivalents in the other lectins. These residues may be the key determinants for oligosaccharide specificity in most Man/Glc and Gal specific lectins. In contrast to the monosaccharide binding site, however, the specific side chains of the residues involved are at least as important as loop lengths, although the latter certainly also plays a role.

4.3. Other specificities

4.3.1. Complex specificity

Some legume lectins are not inhibited by any simple sugar. Therefore, this specificity group has been termed ‘complex’. The crystal structures of two members of the complex type of legume lectins are

known: that of lectin IV from *Griffonia simplicifolia* (GS-IV) and that of PHA-L from *Phaseolus vulgaris*. Only in the case of GS-IV, a structure of a carbohydrate complex has been reported [17]. This structure reveals that the Lewis-b tetrasaccharide (Fucα1–2-Galβ1–3[Fucα1–4]GlcNAcβ1-Me) binds to the lectin with its galactose in the monosaccharide binding site in a way essentially identical to what has been observed for the lectins from coral tree and soybean [15,16]. The main difference between GS-IV and the *E. corallodendron* lectin is again the monosaccharide specificity loop (Val221–Glu225 in GS-IV and Thr216–Glu224 in *E. corallodendron*). This loop hydrogen bonds to the galactose in the combining site of *E. corallodendron* lectin but in GS-IV hydrogen bonds with the α1–4 linked fucose (Fig. 12). In the latter lectin, the corresponding loop (residues Val221 to Glu225) has a rather unusual conformation, including a non-proline *cis*-peptide bond between Val221 and Gly222, which was identified as essential for the specificity of GS-IV.

Legume lectins that can be inhibited by simple monosaccharides show only a low affinity for these saccharides ($K_{dis} = 10^{-2}$–10^{-3} M) (for some recently determined values of the thermodynamic parameters of the association between legume lectins and monosaccharides, see [128–130]). It is therefore not a surprise that the removal of one or two hydrogen bonds abolishes monosaccharide binding. This is exactly what seems to have happened with GS-IV: the hydrogen bonds between the galactose in the monosaccharide binding site with the monosaccharide specificity loop have been removed and are compensated by protein–carbohydrate hydrogen bonds with other residues from the tetrasaccharide. This structure therefore suggests that molecular recognition by the legume lectins with a so-called complex specificity is similar to that occurring in lectins with a defined monosaccharide specificity, but with some of the interactions that normally occur with the monosaccharide removed, so that this monosaccharide cannot bind anymore on its own. It is therefore likely that all lectins of complex specificity will be classifiable into one of the monosaccharide specificity groups when structural information becomes known.

4.3.2. Fucose specific lectins

Several fucose specific legume lectins have been isolated, but only two have been analysed in detail as what concerns the mechanism of carbohydrate specificity: lectin I from gorse (*Ulex europaeus* – UEA-I) [131] and the lectin from *Lotus tetragonolobus* [132,133]. The former lectin has also been crys-

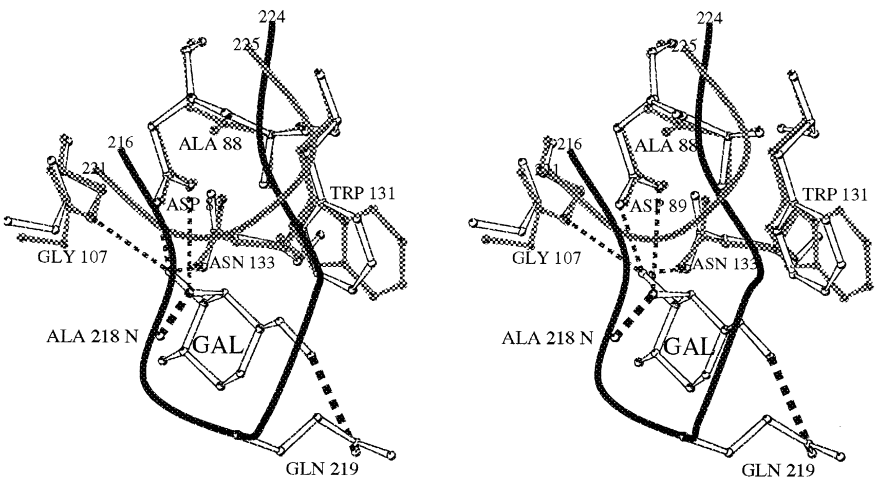


Fig. 12. Superposition of the monosaccharide binding site of *Erythrina corallodendron* lectin and GS-IV. The open ball-and-stick model represents the *Erythrina corallodendron* lectin (including a bound galactose residue), while the grey ball-and-stick model corresponds to the GS-IV structure. The monosaccharide specificity loops are drawn in grey for GS-IV and black for the *Erythrina corallodendron* lectin. Labels correspond to the *Erythrina* structure.

tallised [134], but no structure has been published so far. Some evidence comes from binding studies with deoxy-derivatives of L-Fucα1–2-D-Galβ1–4-D-GlcNAc, the highest affinity ligand known to UEA-I [131]. From these studies, all three hydroxyls of the fucose were identified as involved in key polar interactions, although the effect of deleting the O2 hydroxyl is less dramatic [135], suggesting that O3 and O4 might interact with the conserved Asp–Asn–Gly triad, as is the case for the Gal specific lectins.

Recently, Gohier and co-workers modelled the binding of fucose by UEA-I [136]. This study resulted in several possible orientations for the fucose ring in the binding site, and it was not possible to decide which mode of binding was the most probable one. Therefore, a definite answer on how fucose specificity is achieved will have to await the determination of a crystal structure.

4.3.3. Chitobiose specific lectins

Less information is available on the molecular basis of the specificity of the chitobiose (or GlcNAc) specific legume lectins. Such lectins are found in the seeds of *Ulex europaeus* (UEA-II), *Laburnum alpinum* and *Cytisus* species [137–140] and in the seeds and leaves of *Griffonia simplicifolia* (GS-II) [141]. UEA-II is the only one for which the saccharide specificity has been investigated in more detail [142]. It was found that UEA-II has its highest affinity for L-Fucα1–2-D-Galβ1–4-D-GlcNAcβ1–6-R, the same trisaccharide that is the best known inhibitor of UEA-I, which is fucose specific and shares less than 50% sequence identity with UEA-II. This similarity in oligosaccharide specificity is remarkable in view of their differences in monosaccharide specificity. UEA-I is not inhibited by chitobiose while L-fucose does not at all bind to UEA-II. UEA-II has now been crystallised (our unpublished results) and comparison of this structure with the one of UEA-I will undoubtedly reveal how these two lectins recognise the same oligosaccharide using different epitopes.

Chitobiose specific lectins also weakly interact with some glucosides such as cellobiose, indicating that the *N*-acetyl group is not essential [141,142]. One might therefore speculate that in the case of UEA-II, the monosaccharide binding site is occupied by a GlcNAc or glucose residue in a way similar to the Man/Glc specific lectins, but that the monosaccharide-specificity loop (residues Val223 to Phe229 in UEA-II) does not interact with this GlcNAc or glucose, but rather with one or more residues further on the carbohydrate chain, such as for the complex specificity type of legume lectins. It remains however questionable whether this is compatible with its high affinity for L-Fucα1–2-D-Galβ1–4-D-GlcNAc and L-Fucα1–2-D-Galβ1–4-D-Glc. The Glc/Man specific lectins involve O4 in two key hydrogen bonds, which seems incompatible with any bulky substituent on O4. It should also be noted here that in their original paper describing the fine specificity of UEA-II, Pereira and co-workers showed that these two saccharides have a large area in common with chitobiose that is structurally identical and that is most likely the part of the carbohydrate with which the lectin interacts [142].

Some further evidence comes from a mutagenesis study on lectin II from *Griffonia simplicifolia* [126], which also weakly binds GlcNAc. As expected from similar studies on pea lectin and PHA-L [125,140–144], substitution of the conserved Asp88 and Asn136 in the monosaccharide binding site abolishes carbohydrate binding. On the other hand, the effect of mutating Tyr134 depends upon the nature of the replacement residue: while the mutation Y134F seems to have no effect, Y134D and Y134G eliminates binding of GlcNAc. This can be rationalised when one considers that Tyr134 corresponds to Trp132 in the *Dolichos biflorus* seed lectin, one of the key residues suggested to interact with the *N*-acetyl moiety of GalNAc. Perhaps in the chitobiose or GlcNAc specific lectins, the same amino acids are responsible for interacting with the *N*-acetyl moiety of the GlcNAc that is bound in the monosaccharide binding site. Interestingly, this study also gives evidence for our hypothesis that monosaccharide specificity is determined largely by the length and conformation of the so-called monosaccharide specificity loop (Thr97–Glu102 in concanavalin A, Thr28α–Ala33α in the *Vicieae* lectins and Thr216–Glu224 in the *E. corallodendron* lectin): Point mutations in the corresponding loop of GS-II (Ser225–Ser230) such as L226K and Q227E had no effect on GlcNAc binding, while the deletion of the dipeptide Leu226–Gln227 destroyed GlcNAc binding. In other words, again loop length (and probably conformation) is a more

important determinant than the exact amino acid sequence.

5. Conclusion

Despite the slow start in the seventies and eighties, structural knowledge on legume lectins has boomed in the last five years. Although originally considered a diverse group of proteins, the initial structures of concanavalin A and some *Vicieae* lectins suggested that all of them would have a highly conserved tertiary and quaternary structure. The variation in the quaternary associations now known came as a real surprise. The work by the group of Brewer further suggests that this variation in quaternary structure may be the basis of a form of oligosaccharide specificity previously unrecognised in lectins. The molecular origin of subunit associations in legume lectins, however, is far from clear. Steric hindrance due to bulky covalently attached carbohydrate was initially proposed in the case of GS-IV and the lectin from *E. corallodendron*, but has been revised in the case of GS-IV. Furthermore, PHA-L forms the canonical dimer and is glycosylated at the same positions as the *E. corallodendron* lectin. The lectin from peanut is not glycosylated but nevertheless forms the most remarkable of all tetramers. It is likely that the many legume lectin crystal structures that will be solved in the next years will add to the variety of quaternary structures already available and it is hoped that a synthesis of the available structural information will then lead to a satisfactory explanation of this phenomenon.

A great deal has been learned about the molecular basis of carbohydrate recognition. The principles derived from the early structures have been confirmed numerous times and we have seen glimpses on how different mono- and oligosaccharides can be selected by a conserved scaffold. It appears that a basic set of essential and conserved residues is surrounded by a limited number of variable residues that direct the specificity of the lectin, thus opening interesting perspectives for engineering novel specificities into a legume lectin. Interesting parallels can be drawn between the different lectins, not only for the architecture of the monosaccharide binding sites, but also for oligosaccharide recognition. A further understanding will undoubtedly require a combination of crystallography, site-directed mutagenesis, NMR and modelling. The continuing interest in these proteins nevertheless suggests that further advances may be made in the near future.

Acknowledgements

We wish to thank F. Brewer, A. Imberty, F Casset, M. Etzler, J. Ångstrom, E. Moreno, C. Reynolds, J. Naismith and J. Helliwell for kindly providing coordinates of modelling and crystallographic studies as well as for giving us access to results prior to publication. This work was supported by the Vlaams Interuniversitair Instituut voor Biotechnologie. R. Loris and J. Bouckaert are postdoctoral fellows of the F.W.O. T. Hamelryck received financial support from the Instituut voor Wetenschap en Technologie (I.W.T.).

Note added in proof

The predicted location of the adenine binding site has recently been confirmed by the structure of a complex between the seed lectin from *Dolichos biflorus* and adenine. A crystal structure has also become available of the dimeric stem and leaf lectin (DB58) of *Dolichos biflorus*. This lectin is unusual as it is dimeric but still shows adenine binding activity. In the crystal structure of DB58, a novel type of dimer is observed, in which the adenine binding site of the seed lectin is structurally conserved.

References

[1] N. Sharon, H. Lis, FASEB J. 4 (1990) 3198–3208.
[2] H. Stillmark, Uber Ricin, ein giftiges Ferment aus den Samen von Ricinus comm. L. und einigen anderen Euphorbiaceen, PhD thesis, Dorpat University, Estonia, 1888
[3] K. Fiedler, K. Simons, Cell 77 (1994) 625–626.
[4] K. Fiedler, K. Simons, J. Cell Sci. 109 (1995) 271–276.
[5] C. Itin, A.C. Roche, M. Monsigny, H.P. Hauri, Mol. Biol. Cell 7 (1996) 483–493.
[6] G.M. Edelman, B.A. Cunningham, G.N. Reecke, J.W. Becker, M.W. Waxdal, J.M. Wang, Proc. Natl. Acad. Sci. U.S.A. 69 (1972) 2580–2584.
[7] K.D. Hardman, C.F. Ainsworth, Biochemistry 11 (1972) 4910–4919.

[8] T.W. Hamelryck, M.-H. Dao-Thi, F. Poortmans, M.J. Chrispeels V, L. Wyns, R. Loris, J. Biol. Chem. 271 (1996) 20479–20485.
[9] H. Einspahr, E.H. Parks, K. Suguna, E. Subramanian, F.L. Suddath, J. Biol. Chem. 261 (1986) 16518–16527.
[10] R. Loris, J. Steyaert, D. Maes, J. Lisgarten, R. Pickersgill, L. Wyns, Biochemistry 32 (1993) 8772–8781; Biochemistry 32, 14229
[11] R. Loris, D. Van Overberge, M.-H. Dao-Thi, F. Poortmans, N. Maene, L. Wyns, Proteins Struct. Funct. Genet. 20 (1994) 330–346.
[12] Y. Bourne, C. Abergel, C. Cambillau, M. Frey, P. Rougé, J.-C. Fontecilla-Camps, J. Mol. Biol. 214 (1990) 571–584.
[13] R. Banerjee, S.C. Mande, V. Ganesh, K. Das, V. Dhanaraj, S.K. Mahanta, K. Suguna, A. Surolia, M. Vijiyan, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 227–231.
[14] R. Banerjee, K. Das, R. Ravishankar, K. Suguna, A. Surolia, M. Vijayan, J. Mol. Biol. 259 (1996) 281–296.
[15] A. Dessen, Gupta, S. Dabesan, C.F. Brewer, J.C. Sachettini, Biochemistry, 34 (1995) 4933–4942
[16] B. Shaanan, H. Lis, N. Sharon, Science 254 (1991) 862–866.
[17] L.T.J. Delbaere, M. Vandonselaar, L. Prasad, J.W. Quail, K.S. Wilson, Z. Dauter, J. Mol. Biol. 230 (1993) 950–965.
[18] C. Bompard-Gilles, P. Rousseau, P. Rougé, F. Payan, Structure 4 (1996) 1441–1452.
[19] T.W. Hamelryck, F. Poortmans, A. Goossens, G. Angenon, M. Van Montagu, L. Wyns, R. Loris, J. Biol. Chem. 271 (1996) 32796–32802.
[20] S. Parkin, B. Rupp, H. Hope, Acta Crystallogr. D52 (1996) 1161–1168.
[21] J.R. Helliwell, M. Helliwell, Q. Rev. Biophys. 29 (1996) 227–278.
[22] Y.D. Lobsanov, M.A. Gitt, H. Leffler, S.H. Barondes, J.M. Rini, J. Biol. Chem. 268 (1993) 27034–27038.
[23] D.-I. Liao, G. Kapadia, H. Ahmed, G.R. Vasta, O. Herzberg, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 1428–1432.
[24] J. Emsley, H.E. White, B.P. O'Hara, G. Oliva, N. Srinivasan, I.J. Tickle, T.L. Blundell, M.B. Pepys, S.P. Wood, Nature 367 (1994) 338–345.
[25] S. Crennell, E. Garman, G. Laver, E. Vimr, G. Taylor, Structure 2 (1994) 535–544.
[26] D.D. Leonidas, B.L. Elbert, Z. Zhou, H. Leffler, S.J. Ackerman, K.R. Acharya, Structure 3 (1995) 1379–1393.
[27] P. Argos, T. Tsukihara, M.G. Rossmann, J. Mol. Evol. 15 (1980) 169–179.
[28] G. Chelvanayagam, J. Heringa, P. Argos, J. Mol. Biol. 228 (1992) 220–242.
[29] K.D. Hardman, R.C. Agarwal, M.J. Freiser, J. Mol. Biol. 157 (1982) 69–89.
[30] J. Bouckaert, R. Loris, F. Poortmans, L. Wyns, Proteins Struct. Funct. Genet. 23 (1995) 510–524.
[31] J. Bouckaert, F. Poortmans, L. Wyns, R. Loris, J. Biol. Chem. 271 (1996) 16144–16150.
[32] S.H. Koenig, C.F. Brewer, R.D. Brown, Biochemistry 20 (1978) 4251–4260.
[33] R.D. Brown, S.H. Koenig, C.F. Brewer, Biochemistry 21 (1982) 465–469.
[34] C.F. Brewer, R.D. Brown, S.H. Koenig, J. Biomol. Struct. Dyn. 1 (1983) 961–997.
[35] R.D. Brown, C.F. Brewer, S.H. Koenig, Biochemistry 16 (1977) 3883–3896.
[36] C.F. Brewer, R.D. Brown, S.H. Koenig, Biochemistry 22 (1983) 3691–3702.
[37] T.E. Mirkov, J.M. Wahlstrom, K. Hagiwara, F. Finardi-Filho, Kjemtrup, M.J. Chrispeels, Plant Mol. Biol. 26 (1994) 1103-1113
[38] T.C. Osborn, D.C. Alexander, S.S.M. Sun, C. Cardona, F.A. Bliss, Science 240 (1988) 207–210.
[39] I.J. Goldstein, R.C. Hughes, M. Monsigny, T. Osawa, N. Sharon, Nature 285 (1980) 66.
[40] J. Kocourek, V. Horejsi, Nature 290 (1981) 188.
[41] R. Loris, P.P.G. Stas, L. Wyns, J. Biol. Chem. 269 (1994) 26722–26733.
[42] W.D. McCubbin, C.M. Kay, Biochem. Biophys. Res. Commun. 44 (1971) 101–111.
[43] G.N. Reecke, J.W. Becker, G.M. Edelman, J. Biol. Chem. 250 (1975) 1525–1547.
[44] C. Emmerich, J.R. Helliwell, M. Redshaw, J.H. Naismith, S.J. Harrop, J. Raftery, A.J. Kalb(Gilboa), J. Yariv, Z. Dauter, K.S. Wilson, Acta Crystallogr. D50 (1994) 749–756.
[45] J.H. Naismith, J. Habash, S. Harrop, J.R. Helliwell, W.N. Hunter, T.C.M. Wan, S. Weisgerber, A.J. Kalb(Gilboa), J. Yariv, Acta Crystallogr. D49 (1993) 561–571.
[46] J.H. Naismith, C. Emmerich, J. Habash, S.J. Harrop, J.R. Helliwell, W.N. Hunter, J. Raftery, J. Kalb(Gilboa), J. Yariv, Acta Crystallogr. D50 (1994) 847–858.
[47] S. Weisgerber, J.R. Helliwell, J. Chem. Soc. Faraday Trans. 89 (1993) 2667–2675.
[48] F. Bouckaert, L. Wyns, R. Loris, in: E. Van Driessche, T.C. Bøg-Hansen (Eds.), Lectins: Biology, Biochemistry and Clinical Biochemistry, vol. 11, 1996, pp. 50–60
[49] R. Loris, D. Maes, F. Poortmans, L. Wyns, J. Bouckaert, J. Biol. Chem. 271 (1996) 30614–30618.
[50] P.N. Kanellopoulos, P.A. Tucker, K. Pavlou, B. Agianian, S.J. Hamodrakas, J. Struct. Biol. 117 (1996) 16–23.
[51] J. Sanz-Aparicio, J. Hermoso, T.B. Grangeiro, J.J. Calvete, B.S. Cavada, FEBS Lett. 405 (1997) 114–118.
[52] D.C.H. Yang, W.E. Gall, G.M. Edelman, J. Biol. Chem. 249 (1974) 7018–7023.
[53] G.M. Edelman, J.L. Wang, J. Biol. Chem. 253 (1978) 3016–3022.
[54] D.D. Roberts, I.J. Goldstein, J. Biol. Chem. 257 (1982) 11274–11277.
[55] D.D. Roberts, I.J. Goldstein, J. Biol. Chem. 258 (1983) 13820–13824.
[56] M.J. Maliarik, I.J. Goldstein, J. Biol. Chem. 263 (1988) 11274–11279.
[57] D.D. Roberts, P. Arjunan, L.B. Townsend, I.J. Goldstein, Phytochemistry 25 (1986) 589–593.
[58] C.V. Gegg, D.D. Roberts, I.H. Segel, M.E. Etzler, Biochemistry 31 (1992) 6938–6942.

[59] K.D. Puri, A. Surolia, J. Biol. Chem. 269 (1994) 30917–30926.
[60] T.J.V. Higgins, P.M. Chandler, G. Zurawaski, S.C. Button, D. Spencer, J. Biol. Chem. 258 (1983) 9544–9549.
[61] N.M. Young, D.C. Watson, P. Thibault, Glycoconjugate J. 13 (1996) 575–583.
[62] D.K. Mandal, E. Nieves, L. Bhattacharyya, G.A. Orr, J. Roboz, Q.-T. Yu, C.F. Brewer, Eur. J. Biochem. 221 (1994) 547–553.
[63] M.E. Etzler, Biochemistry 33 (1994) 9778–9783.
[64] N.M. Young, D.C. Watson, M. Yaguchi, R. Adar, R. Arango, E. Rodriguez-Arango, N. Sharon, P.K.S. Blay, P. Thibault, J. Biol. Chem. 270 (1995) 1563–1570.
[65] W. Min, A.J. Dunn, D.H. Jones, EMBO J. 11 (1992) 1303–1307.
[66] P.S. Sheldon, D.J. Bowles, EMBO J. 11 (1992) 1297–1301.
[67] D.M. Carrington, A. Auffret, D.E. Hanke, Nature 313 (1985) 64–67.
[68] D.J. Bowles, S.E. Marcus, D.J. Pappin, J.B. Findlay, E. Eloipoulos, P.R. Maycox, J. Burgess, J. Cell. Biol. 102 (1986) 1284–1297.
[69] E.M. Herman, L.M. Shannon, M.J. Chrispeels, Planta 165 (1985) 23–29.
[70] T.V. Brennan, S. Clarke, Biochem. Biophys. Res. Commun. 193 (1993) 1031–1037.
[71] Y. Abe, K. Shirane, H. Yokoshawa, H. Matsushita, M. Mitta, I. Kato, S.-I. Ishii, J. Biol. Chem. 268 (1993) 3525–3529.
[72] W. Min, D.H. Jones, Nat. Struct. Biol. 1 (1994) 502–504.
[73] Q. Chao, C. Casalongue, J.M. Quinn, M.E. Etzler, Arch. Biochem. Biophys. 313 (1994) 346–350.
[74] C.F. Brewer, L. Bhattacharyya, Biochem. Biophys. Res. Commun. 137 (1986) 670–674.
[75] L. Bhattacharyya, M.I. Kahn, J. Frant, C.F. Brewer, Biochemistry 27 (1988) 8762–8767.
[76] L. Bhattacharyya, M.I. Khan, C.F. Brewer, J. Biol. Chem. 264 (1989) 11543–11545.
[77] D.K. Mandal, C.F. Brewer, Biochemistry 31 (1992) 12602–12609.
[78] D.K. Mandal, C.F. Brewer, Biochemistry 31 (1994) 12602–12609.
[79] C.F. Brewer, Chemtracts: Biochemistry and Molecular Biology 6 (1996) 165–179.
[80] C.F. Brewer, Lectins: Biology, Biochemistry, Clin. Biochem. 11 (1996) 10–16.
[81] J.B. Sumner, S.F. Howell, J. Biol. Chem. 115 (1936) 583–588.
[82] J.L. Wang, B.A. Cunningham, M.J. Waxdal, G.M. Edelman, J. Biol. Chem. 250 (1975) 1490–1502.
[83] B.A. Cunningham, J.L. Wang, M.J. Waxdal, G.M. Edelman, J. Biol. Chem. 250 (1975) 1503–1512.
[84] Z. Derewenda, J. Yariv, J.R. Helliwell, A.J. Kalb(Gilboa), E.J. Dodson, M.Z. Papiz, T. Wan, J. Campbell, EMBO J. 8 (1989) 2189–2193.
[85] B.A. Williams, M.C. Chervenak, E.J. Toone, J. Biol. Chem. 267 (1992) 22907–22911.
[86] D.K. Mandal, N. Kishore, C.F. Brewer, Biochemistry 33 (1994) 1149–1156.
[87] M.C. Chervenak, E.J. Toone, Biochemistry 34 (1995) 5685–5695.
[88] P.N. Kanellopoulos, K. Pavlou, A. Perrakis, B. Agianian, C.E. Vorgias, C. Mavrommatis, M. Soufi, P.A. Tucker, S.J. Hamodrakas, J. Struct. Biol. 116 (1996) 345–355.
[89] S.J. Hanodrakas, P.N. Kanellopoulos, K. Pavlou, P.A. Tucker, J. Struct. Biol. 118 (1997) 23–30.
[90] J.U. Baezinger, D. Fiete, J. Biol. Chem. 254 (1979) 2400–2407.
[91] N. Debray, D. Decout, G. Strecker, G. Spik, J. Montreuil, Eur. J. Biochem. 117 (1981) 41–55.
[92] J.H. Naismith, R.A. Field, J. Biol. Chem. 271 (1996) 972–976.
[93] D.K. Mandal, L. Bhattacharyya, S.H. Koenig, R.D. Brown, S. Oscarson, C.F. Brewer, Biochemistry 33 (1994) 1157–1162.
[94] D. Gupta, T.K. Dam, S. Oscarson, C.F. Brewer, J. Biol. Chem., in press, 1997
[95] V.S. Reddy, V.S.R. Rao, Int. J. Biol. Macromol. 14 (1991) 185–192.
[96] G.N. Reecke, J.W. Becker, Science 234 (1986) 1108–1111.
[97] Y. Bourne, A. Roussel, M. Frey, P. Rougé, J. Fontecilla-Camps, C. Cambillau, Proteins 8 (1990) 365–376.
[98] J.M. Rini, K.D. Hardman, H. Einspahr, F.L. Suddath, J.P. Carver, J. Biol. Chem. 268 (1993) 10126–10132.
[99] F. Casset, T. Hamelryck, R. Loris, J.-R. Brisson, C. Teller, M.-H. Dao-Thi, L. Wyns, F. Poortmans, S. Pérez, A. Imberty, J. Biol. Chem. 270 (1995) 25619–25629.
[100] Y. Bourne, A. Ayouba, P. Rougé, C. Cambillau, J. Biol. Chem. 269 (1994) 9429–9435.
[101] R. Loris, F. Casset, J. Bouckaert, J. Pletinckx, M.-H. Dao-Thi, F. Poortmans, A. Imberty, S. Pérez, L. Wyns, Glycoconjugate J. 11 (1994) 507–517.
[102] D. Loganathan, S.E. Osborne, G.D. Glick, I.J. Goldstein, Arch. Biochem. Biophys. 299 (1992) 268–274.
[103] Y. Bourne, P. Rougé, C. Cambillau, J. Biol. Chem. 265 (1990) 18161–18165.
[104] Y. Bourne, P. Rougé, C. Cambillau, J. Biol. Chem. 267 (1992) 197–203.
[105] A. Imberty, Y. Bourne, C. Cambillau, P. Rougé, S. Pérez, Adv. Biophys. Chem. 3 (1993) 71–118.
[106] K. Kornfeld, M.L. Teitman, R. Kornfeld, J. Biol. Chem. 256 (1981) 6633–6640.
[107] Y. Bourne, J. Mazurier, D. Legrand, P. Rougé, J. Montreuil, G. Spik, C. Cambillau, Structure 2 (1994) 209–219.
[108] M. Ueno, H. Ogawa, I. Matsumoto, N. Seno, J. Biol. Chem. 266 (1991) 3146–3153.
[109] T. Animashaun, R.C. Hughes, J. Biol. Chem. 264 (1989) 4657–4663.
[110] D. Chawla, T. Animashaun, R.C. Hughes, FEBS Lett. 298 (1992) 291–296.
[111] S. Shibata, I.J. Goldstein, D.A. Baker, J. Biol. Chem. 257 (1982) 9324–9329.
[112] N.K. Vyas, M.N. Vyas, F.A. Quiocho, J. Biol. Chem. 226 (1991) 5226–5237.
[113] F.A. Quiocho, D.K. Wilson, N.K. Vyas, Nature 340 (1984) 404–407.

[114] A. Imberty, F. Casset, C.V. Gegg, M.E. Etzler, S. Pérez, Glycoconjugate J. 11 (1994) 400–413.
[115] F. Casset, A. Imberty, S. Pérez, M.E. Etzler, H. Paulsen, T. Peters, Eur. J. Biochem. 244 (1997) 242–250.
[116] F. Casset, T. Peters, M. Etzler, E. Korchagina, N. Nifantev, S. Pérez, A. Imberty, Eur. J. Biochem. 239 (1996) 710–719.
[117] E. Moreno, S. Teneberg, R. Adar, N. Sharon, K.-A. Karlsson, J. Ångström, Biochemistry 36 (1997) 4429–4437.
[118] W.R. Kinzy, D.M. Belenky, F.G. Loontiens, N. Sharon, Glycoconjugate J. 9 (1992) 225–227.
[119] H. Lis, F.J. Joubert, N. Sharon, Phytochemistry 24 (1985) 2803–2809.
[120] M.J. Swami, D. Gupta, S.K. Mahanta, A. Surolia, Agglutinin, Carbohydrate Res. 213 (1991) 59–67.
[121] R. Arango, E. Rodriguez-Arango, R. Adar, D. Belenky, F.G. Loontiens, S. Rozenblatt, N. Sharon, FEBS Lett. 330 (1993) 133–136.
[122] E.T. Jordan, I.J. Goldstein, Eur. J. Biochem. 230 (1995) 958–964.
[123] N.M. Young, R.P. Oomen, J. Mol. Biol. 228 (1992) 924–934.
[124] V. Sharma, A. Surolia, J. Mol. Biol. 267 (1997) 433–445.
[125] R. Adar, N. Sharon, Eur. J. Biochem. 239 (1996) 668–674.
[126] K. Zhu, R.A. Bressan, P.M. Hasegawa, L.M. Murdock, FEBS Lett. 390 (1996) 271–274.
[127] M. Nishiguchi, K. Yoshida, T. Sumizono, K. Tazaki, FEBS Lett. 403 (1997) 294–298.
[128] S. Acharya, S.R. Patanjali, S.U. Sajjan, B. Gopalakrishnan, A. Surolia, J. Biol. Chem. 265 (1990) 11586–11594.
[129] F.P. Schwartz, K.D. Puri, R.G. Bhat, A. Surolia, J. Biol. Chem. 268 (1993) 7668–7677.
[130] A. Surolia, N. Sharon, F.P. Schwartz, J. Biol. Chem. 271 (1996) 17697–17703.
[131] M.E.A. Pereira, E.C. Kisailus, F. Gruezo, E.A. Kabat, Arch. Biochem. Biophys. 185 (1978) 108–115.
[132] J. Yariv, A.J. Kalb, E. Katchalski, Nature 215 (1967) 890–891.
[133] M.E.A. Pereira, E.A. Kabat, Biochemistry 13 (1974) 3184–3192.
[134] M. Vandonselaar, L.T.J. Delbaere, J. Mol. Biol. 243 (1994) 345–346.
[135] M.-H. Du, U. Spohr, R.U. Lemieux, Glycoconjugate J. 11 (1994) 443–461.
[136] A. Gohier, J.F. Espinosa, J. Jimenez-Barbero, P.A. Carrupt, S. Pérez, A. Imberty, J. Mol. Graph, in Press, 1997
[137] I. Matsumoto, T. Osawa, Arch. Biochem. Biophys. 140 (1970) 491–585.
[138] Y. Konami, K. Yamamoto, T. Tsuji, I. Matsumoto, T. Osawa, Hoppe-Seyler's Z. Physiol. Chem. 364 (1983) 397–405.
[139] Y. Konami, K. Yamamoto, T. Osawa, The primary structures of two types of the Ulex europeus seed lectin, J. Biochem. 109 (1991) 650–658.
[140] Y. Konami, K. Yamamoto, T. Osawa, Biol. Chem. Hoppe-Seyler 372 (1991) 103–111.
[141] P.N.S. Iyer, K.D. Wilkinson, I.J. Goldstein, Arch. Biochem. Biophys. 177 (1976) 330–333.
[142] M.E.A. Pereira, F. Gruezo, E.A. Kabat, Arch. Biochem. Biophys. 194 (1979) 511–525.
[143] R.R. Van Eijsden, F.J. Hoedemaeker, C.L. Diaz, B.J.J. Lugtenberg, B.S. De Pater, J.W. Kijne, Plant Mol. Biol. 20 (1992) 1049–1058.
[144] T.E. Mirkov, M.J. Chrispeels, Glycobiology 3 (1993) 581–587.
[145] J. Moreno, M.J. Chrispeels, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 7885–7889.
[146] D.N. Figlas, H.R. Arias, A. Fernandez, D.M. Alperin, Dramatic saccharide-mediated protection of chaotropic induced deactivation of concanavalin A, Arch. Biochem. Biophys. 340 (1997) 154–158.
[147] E. Osinaga, D. Tello, C. Batthyany, M. Bianchet, R. Durán, C. Cervenansky, L. Canoin, P. Paroutaud, A. Roseto, P. Alzari, Preliminary crystallographic analysis and amino acid sequence of isolectin B4 from *Vicia villosa*, Abstracts of the 16th Interlec Meeting, Toulouse France, 1995
[148] J.K. Scott, D. Loganathan, R.B. Easley, X. Gong, I.J. Goldstein, A family of concanavalin A-binding peptides from a hexapeptide epitope library, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 5398–5402.
[149] K.R. Oldenburg, D. Loganathan, I.J. Goldstein, P.G. Schultz, M.A. Gallop, Peptide ligands for a sugar-binding protein isolated from a random peptide library, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 5393–5397.
[150] K.J. Kaur, S. Khurana, D.M. Salunke, Topological analysis of the functional mimicri between a peptide and a carbohydrate moiety, J. Biol. Chem. 272 (1997) 5539–5543.
[151] A.R. Shickman, N.S. Greenspan, M.W. Cunningham, Cytokeratin peptide SFGSGFGGGY mimics *N*-acetyl β-D-glucosamine in reaction wih antibodies and lectins, and induces in vivo anti-carbohydrate antibody response, J. Immunol. 153 (1994) 5593–5606.
[152] R. Ravishankar, M. Ravindran, K. Suguna, A. Surolia, M. Vijayan, Crystal structure of the peanutlectin – T-antigen complex. Carbohydrate specificity generated by water bridges, Curr. Sci. 72 (1997) 855–861.
[153] N. Sharon, Lectin-carbohydrate complexes of plants and animals: an atomic view, Trends Biochem. Sci. 18 (1993) 221–226.
[154] R.R. Van Eijsden, B.S. De Paters, J.W. Kijne, Mutational analysis of the sugar binding site of pea lectin, Glycoconjugate J. 11 (1994) 3775–3780.
[155] T. Sokolowski, T. Peters, S. Pérez, A. Imberty, J. Mol. Graph 15 (1997) 37–42.
[156] P.J. Kraulis, J. Appl. Cryst. 24 (1993) 946–950.
[157] A. Nicholls, K.A. Sharp, B. Honig, Proteins Struct. Funct. Genet. 11 (1991) 281–296.