Novel structures of plant lectins and their complexes with carbohydrates
Julie Bouckaert, Thomas Hamelryck, Lode Wyns and Remy Loris*

Several novel structures of legume lectins have led to a thorough understanding of monosaccharide and oligosaccharide specificity, to the determination of novel and surprising quaternary structures and, most importantly, to the structural identification of the binding site for adenine and plant hormones. This deepening of our understanding of the structure/function relationships among the legume lectins is paralleled by advances in two other plant lectin families — the monocot lectins and the jacobin family. As the number of available crystal structures increases, more parallels between plant and animal lectins become apparent.

Introduction

The past year has witnessed the determination of a wide variety of novel structures of plant lectins, most of them legume lectins, in complex with carbohydrate ligands. This review will, therefore, put emphasis on recent knowledge gained in the domain of the legume lectins, but major results from some other plant lectin families will be discussed briefly as well. Finally, we will discuss some striking similarities that have emerged among the quaternary structures of plant and animal lectins.

Legume lectin–carbohydrate complexes

A comprehensive comparative study by Sharma and Surolia published in 1997 [1] aimed to rationalize the structural basis of monosaccharide specificity in the legume lectins. Since then, several crystal structures of novel Gal/GalNAc-binding lectins have appeared, against which these views could be tested.

Abbreviations

Abbreviations

Con A concanavalin A
DBL Dolichos biflorus lectin
EcorL Erythrina corallodendron lectin
Fuc fucose
Gal galactose
GalNAc N-acetylgalactosamine
Glc glucose
GlcNAc N-acetylglucosamine
Man mannose
MPA Majus pomifera agglutinin
PNA peanut agglutinin
PSP porcine seminal plasma spermadhesin
WBA winged bean agglutinin

Complexes of the Dolichos biflorus seed lectin (DBL) (Figure 1a) with the blood group A trisaccharide (GalNAc(α1-3)Gal(α1-2)Gal) and the Forssman antigen (GalNAc(α1-3)Gal) convincingly showed that Leu127 plays an essential role both in the low affinity of DBL for Gal and in its specificity for the Forssman antigen. Leu127 is substituted by an aromatic residue in most other legume lectins. Substitution of this aromatic residue by a shorter, aliphatic leucine or valine residue has been shown to abolish carbohydrate binding in the Gal-specific lectins from Erythrina corallodendron (EcorL) [3] and Robinia pseudoacacia [4]. Indeed, the mutation L127F increases the affinity of DBL for Gal and GalNAc. The reduction in the contact surface area as a result of the presence of Leu127 is compensated by additional hydrophobic interactions between the lectin and the methyl group of the N-acetyl moiety of GalNAc and by an additional hydrogen bond between the N-acetyl group and Glu103. Such a mechanism for the specific recognition of N-acetylated sugars, whereby the lack of stacking of a large aromatic sidechain against the sugar ring is replaced by compensating interactions with the N-acetyl group, seems to be a recurring feature in protein–carbohydrate recognition and was recently observed in the complex of tachylectin-2 with GlcNAc [5••].

This contrasts with the situation in EcorL, which has a similar affinity for Gal and GalNAc. In EcorL, (Figure 1b), fewer contacts are seen between the lectin and the N-acetyl group of GalNAc and, specifically, the hydrogen bond with the conserved Gly107 residue (Gly103 in DBL) is missing [6]. This is a direct result of the slightly different orientation of GalNAc in the binding site of EcorL, compared with that in DBL. This is thought to be a consequence of the substitution of Glu129 in EcorL by Ser215 in DBL [2••].

A third variation exists in peanut agglutinin (PNA) (Figure 1c), which, in contrast to DBL, strongly prefers Gal to GalNAc. In PNA, the substitution of Glu129 by aspartic acid converts this strictly Gal-specific lectin to one that binds both Gal and GalNAc [7]. In PNA, Leu127 of DBL is replaced by the more common Tyr125 residue, whereas Glu129 of PNA has no structural counterpart in DBL.

Finally, yet another variation is seen in the complex between Gal and the basic lectin from winged bean, winged bean agglutinin (WBA) (Figure 1d) [8]. WBA has an unusually long, variable ‘monosaccharide specificity loop’ that is argued to be essential both for the strict preference of WBA for Gal linkages and for its enhanced affinity for galactoses bearing substitutions (such as an N-acetyl group) on C2.
The high affinity of DBL for the disaccharide GalNAc(\(\alpha_1-3\))GalNAc was shown to be as a result of the shielding of a hydrophobic patch on the lectin surface by the N-acetyl group of the reducing GalNAc that is positioned outside of the monosaccharide-binding site [2••]. This is broadly in agreement with earlier modeling and NMR studies [9,10]. Shielding of hydrophobic surface has also been invoked as the prime reason for the four-fold increase in the binding affinity of Man(\(\alpha_1-3\))Man for concanavalin A (Con A) compared with \(\alpha\)-D-mannose [11•]. In this complex, the reducing mannose also does not form specific hydrogen bonds to the protein.

Several more reports have appeared on Con A [11•,12,13,14•] and the related lectin from Dioscorea grandiflora [15] that focus on the relationship between structure and thermodynamics. This relationship has proven to be a difficult one to rationalize, as, in most complexes, a trade-off is observed between van der Waals and hydrogen-bond interactions on the one hand and conformational strain in the glycosidic linkages and the destabilization of binding in the monosaccharide-binding site on the other hand [11•,13,14•]. Thus, it is not surprising that computational studies aimed at predicting and understanding experimental thermodynamic parameters show only poor success.
rates, even when applied to the very simple system of Man versus Glc binding to Con A [12].

Novel quaternary structures of legume lectins

The legume lectin family is known to exhibit a number of different quaternary structures. Several recent reports have now further extended this variety. WBA [8] is the second lectin to have been found to adopt the EcorL dimer structure. In contrast to EcorL, glycosylation cannot be invoked to explain the noncanonical association of the WBA monomers. In their paper, Prabu et al. [8] present a detailed discussion of noncanonical association in legume lectin dimers.

Most interesting are the structures of two related lectins from *Dolichos biflorus* (DBL, from the seeds, and DB58, from the stem and leaves) (Figure 2) [2••]. These lectins each consist of an equimolar mixture of full-length and C-terminally truncated subunits. In the crystal structures of DBL and DB58, these C termini were shown to adopt an α-helical conformation and their α helices are packed between the back β sheets of two opposing monomers. In the case of DB58, this leads to the formation of a novel kind of dimer. The C-terminal helices not only mediate quaternary structure formation, but also play an important role in establishing the adenine/hormone-binding site found in many legume lectins.

GlcNAc-specific lectin II from *Ulex europeaus* has recently been crystallized and a preliminary structure determination reveals a tetramer resembling PHA-L (phytohaemagglutinin-L, from *Phaseolus vulgaris*), SBA (soybean agglutinin) and DBL [16]. In contrast to DBL, no central helix is present between the two back β sheets. Rather, the tetrameric arrangement is stabilized by interprotomer disulfide bridges (R Loris, unpublished data). Electron microscopy and X-ray diffraction studies at low resolution suggested a novel, tetrameric arrangement for fucose-specific isolectin A from *Lotus tetragonolobus* [17•]. This work constitutes the first detailed characterization of a type-II two-dimensional cross-linked lattice between a lectin and a bivalent carbohydrate, the oncofetal dimeric Lewis* α* antigen. Fitting the canonical legume lectin dimer to the 19 Å electron diffraction data resulted in a tetramer...
in which the two canonical dimers are rotated 20° relative to their mutual orientations in the Con A tetramer.

Recently, an attempt was made to rationalize quaternary structure formation in legume lectins in terms of buried hydrophobic surface area, shape complementarity and calculated interaction energy [18]. Neither of these parameters allows a completely unambiguous discrimination between arrangements that are experimentally observed and those that are not; however, globally, each parameter tends to favor the observed arrangements.

Structural link between plant and animal lectins

Legume lectins, galectins, pentraxins and spermadhesins are all carbohydrate-binding proteins that share a related monomeric fold. In all four families, a variety of quaternary structures is observed, giving rise to different spacings among the binding sites of different monomers (for the most recent additions to the universe of lectin quaternary structures, see [2••,19–22]). The spacing and orientation of the sugar-binding sites of a lectin become important when binding multivalent epitopes [23,24].

That evolutionarily related proteins adopt different quaternary structures is not unusual as such. What is unusual is that, among these lectins, variation in quaternary structure is possible despite a large degree of conservation in amino acid sequence and in monomeric structure. Furthermore, some of the individual quaternary structures observed for evolutionarily unrelated legume lectins, galectins and spermadhesins are strikingly similar (Figure 3). The first crystal structures of the galectin dimer showed a striking resemblance to the canonical legume lectin dimer [25,26]. Similarly, in the crystal structure of heterodimeric PSP-I–PSP-II (porcine seminal plasma spermadhesin), two PSP-I monomers associate to form a structure that resembles a canonical legume lectin dimer [20]. Given that sheet extension is an often-used mechanism for dimer formation in many unrelated proteins, this observation could still be attributed to chance. On the other hand, the similarities in subunit association among the spermadhesins and GS-IV (Griffonia simplicifolia lectin IV) or PNA are amazing to say the least. The PSP-I–PSP-II spermadhesin heterodimer strongly resembles the dimer formed by the legume lectin GS-IV [20]. Together with the above-mentioned crystallographic association of the PSP-I subunits, a PNA-like tetramer is formed. PNA was the first protein to be shown to adopt an ‘open’ quaternary structure [27]. Such structures were, up to a few years ago, considered ‘forbidden’ by nature.

Perhaps the lectin folds were chosen during evolution exactly because they allow the easy evolution of quaternary structures, in order that they can adapt easily to the different spacing of carbohydrate ligands presented on, for example, cell surfaces. As a result of the intrinsic flexibility of oligosaccharides, high-affinity binding poses a specific problem for carbohydrate-binding proteins. Thus, it is possible that multivalency is nature’s solution to this problem and that the rapid evolution of quaternary structure is a necessary consequence.

Monocot lectins

Three recent reports provide insights into the structural basis of the oligosaccharide specificities of the monocot lectins, as well as into quaternary structure formation [28•,29•,30]. The dimeric lectin from garlic (Allium sativum) is the second monocot lectin for which a detailed structural analysis has become available [28•]. Seven mannoses are found bound to this dimer: one bound to each subdomain of the threefold pseudo-symmetrical monomer
and one further mannose bound to a high-affinity subsite that is thought to be relevant to the recognition of a high mannose structure (Figure 4a). The differences in carbohydrate specificity between dimeric galactin and tetrameric snowdrop lectin appear to be a direct consequence of their different quaternary structures. Thus, the monocot lectins are the first example of a protein family that uses variation in subunit association as a mode of achieving oligosaccharide specificity.

The crystal structure of daffodil lectin in complex with Mastr-3-Man also reveals an additional binding site that is distinct from the three carbohydrate-recognition domains symmetrically arranged on each monomer [29•]. This fourth binding site is located in a different position to that of galactin (Figure 4b) and was also proposed to play a role in the recognition of larger oligosaccharide ligands, as is further supported by a comparison with the snowdrop lectin-pentasaccharide complex [31•](Figure 4c). For the ‘primary’ binding sites, the disaccharide can bind with either its reducing or nonreducing end in the monosaccharide-binding site, a situation similar to the Manα1-2-Manβ1–Con A complex.

The jacalin family

The crystal structure determination of a second member of the jacalin family, Masca pinnata agglutinin (MPA) [32], resulted in the first view of how the T-antigen disaccharide (Galβ1-3GalNAcβ) is specifically recognized by this lectin family. Interestingly, the bulk of the interactions between the carbohydrate and the lectin occur via the reducing end of GalNAc, whereas the nonreducing Gal only interacts with the protein via a small number of van der Waals contacts and a single water bridge. Thus, the preference of MPA for Galβ1-3GalNAc over GalNAc is explained essentially by the occurrence of the shielding of exposed hydrophobic surface, as has also been observed recently with GalNAcα1-3GalNAc binding to DBL [22•] and Masca-3-Man binding to Con A [11•].

A plausible model for discrimination between Man and Gal by KM+•, a mannose-specific member of the jacalin family, has been presented based on homology modeling [33•]. A key role in Man and Gal distinction by KM+ may be played by Asp141. In contrast to jacalin and MPA, all known mannose-specific lectins from the same family are single-chain lectins that are not post-translationally cleaved into an α and β chain. In KM+, Gly14 of the link-er peptide between the virtual α and β chains forces Asp141 into a different rotamer conformation than observed for the equivalent Asp125p in the jacalin-Gal complex. In this conformation, Asp141 could make a hydrogen bond with an equatorial oxygen on C4, as found in Man and Glu, but not with an axial O4, as in Gal.

Conclusions

The legume lectins continue to profile themselves as the model system of choice to study the fundamentals of protein-carbohydrate interaction. The analogies one continues to find among these plant lectins and several families of animal lectins justify this choice. A full understanding of the structural basis of carbohydrate specificity and quaternary structure formation in the legume lectin family requires detailed studies on members belonging to other specificity groups (e.g. Fuc or GlcNAc), as well as detailed mutagenesis and calorimetric analysis. Several groups are indeed working on this aim.

Other plant lectin families have now also been structurally characterized in more detail. While even a year ago, only a single member of the jacalin family and the monocot lectins was known, relatives with different specificities can now be compared. Several crystalization reports suggest that our understanding will also grow rapidly in the coming years.

Acknowledgements

We thank P. Roberts (NBS, University, UK) and C. Reynolds (Liverpool John Moores, UK) for discussions and for providing data prior to publication. L. Greune and R. Gurrer (Uni Paolo, Bari) kindly provided the coordinates of the model of KM+•. R. Loris and J. Bouckaert are acknowledged for their generous help. This work would be supported by the Vlaams Interuniversitair Fonds voor de Biologische Wetenschappen. We thank Maria Vanderkerckh and Yves Geurnes for excellent technical assistance.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:
• • of special interest
** • • of outstanding interest

Novel plant lectin structures


11. Prabu MM, Sankaranarayanan R, Puri KD, Sharma V, Surolia A, Prabhu MM, Sankaranarayanan R, Puri KD, Sharma V, Surolia A. The crystal structures of concanavalin A complexed with Man(1-3)Man(1-2)Man(1-6)Man(1-2)Man(1-6)Man(1-2)Man and Man(1-3)Man(1-4)GlcNAc and Man(1-3)Man(1-2)Man(1-3)GlcNAc are discussed in view of recently published thermodynamic data. The large change in heat capacity upon binding Man(1-3)Man(1-2)Man(1-3)GlcNAc agrees well with the binding of the reducing mannosyl residue to a hydrophobic surface.


19. The first detailed structural characterization of a type II two-dimensional cross-link provides a better model for the cross-linking processes that occur at cell surfaces than the better studied three-dimensional cross-links.


24. The first report to explain mannose and galactose discrimination in the jacalin family. An analysis is also given of conserved structural determinants of the β-prism fold.
