doi:10.1006/jmbi.2001.4639 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2001) 309, 193–201

JMB

Weak Protein-Protein Interactions in Lectins: The Crystal Structure of a Vegetative Lectin from the Legume *Dolichos biflorus*

Lieven Buts[1]*, Minh-Hoa Dao-Thi[1], Remy Loris[1], Lode Wyns[1] Marilynn Etzler[2] and Thomas Hamelryck[1]

[1]*ULTR-Ultrastructure Department, Vrije Universiteit Brussel, Paardenstraat 65 B-1640, Sint-Genesius-Rode Belgium*

[2]*Section of Molecular and Cellular Biology, University of California, Davis CA 95616, USA*

**Corresponding author*

The legume lectins are widely used as a model system for studying protein-carbohydrate and protein-protein interactions. They exhibit a fascinating quaternary structure variation, which becomes important when they interact with multivalent glycoconjugates, for instance those on cell surfaces. Recently, it has become clear that certain lectins form weakly associated oligomers. This phenomenon may play a role in the regulation of receptor crosslinking and subsequent signal transduction. The crystal structure of DB58, a dimeric lectin from the legume *Dolichos biflorus* reveals a separate dimer of a previously unobserved type, in addition to a tetramer consisting of two such dimers. This tetramer resembles that formed by DBL, the seed lectin from the same plant. A single amino acid substitution in DB58 affects the conformation and flexibility of a loop in the canonical dimer interface. This disrupts the formation of a stable DBL-like tetramer in solution, but does not prohibit its formation in suitable conditions, which greatly increases the possibilities for the crosslinking of multivalent ligands. The non-canonical DB58 dimer has a buried symmetrical α helix, which can be present in the crystal in either of two antiparallel orientations. Two existing structures and datasets for lectins with similar quaternary structures were reconsidered. A central α helix could be observed in the soybean lectin, but not in the leucoagglutinating lectin from *Phaseolus vulgaris*. The relative position and orientation of the carbohydrate-binding sites in the DB58 dimer may affect its ability to crosslink mulitivalent ligands, compared to the other legume lectin dimers.

Keywords: legume lectins; *Dolichos biflorus*; quaternary structure; X-ray crystallography

Introduction

Lectins are a structurally diverse class of proteins of non-immune origin that bind carbohydrates in a reversible fashion and do not exhibit enzymatic activity towards their ligands.[1] Lectins can be found in plants, animals, bacteria and viruses, and are involved in countless biological recognition functions. The legume lectin family has been studied for several decades using biochemical and biophysical techniques, and is considered a model system for protein-carbohydrate interactions.[2] These dimeric or tetrameric proteins are found in the seeds and in the vegetative tissues of most leguminous plants. The monomers exhibit high levels of sequential and structural identity. At present, the crystal structures of 21 native or sugar-

Abbreviations used: ConA, Jack bean (*Canavalia ensiformis*) lectin; DBL, *Dolichos biflorus* (now *Macrotyloma uniflorum*) seed lectin; DB58, 58 kDa vegetative lectin from *Dolichos biflorus*; EcorL, *Erythrina corallodendron* lectin; FRIL, Flt3 receptor interacting lectin from *Dolichos lablab*; GS-IV, *Griffonia simplicifolia* isolectin IV; MAH, *Maackia amurensis* hemagglutinin; MAL, *Maackia amurensis* leucoagglutinin; NCS, non-crystallographic symmetry; PEG, polyethylene glycol; PHA-E, phytohaemagglutinin-E (*Phaseolus vulgaris*); PHA-L, phytohaemagglutinin-L (*Phaseolus vulgaris*); PNA, peanut (*Arachis hypogaea*) agglutinin; SBA, soybean (*Glycine max*) agglutinin; UEA-I and UEA-II, isolectins I and II from *Ulex europaeus*; WBA-I, basic winged bean (*Psophocarpus tetragonolobus*) agglutinin.

E-mail address of the corresponding author: lieven@ultr.vub.ac.be

0022-2836/01/010193–9 $35.00/0

complexed legume lectins have been determined. For a complete overview, consult the 3D Lectin Database, available on-line†.

Despite the conserved structure of the legume lectin monomer, a surprising variety of quaternary structures is observed. This variation has significant functional implications for the binding of multivalent sugar ligands. The symmetry properties of the protein oligomer and the multivalent ligand both affect the structure of homogeneous crosslinked lattices, generating greater specificity than can be achieved at the level of individual monomers.[3] Specific and homogeneous two or three-dimensional crosslinked complexes can be formed even in the presence of mixtures of different lectins and carbohydrates. This additional level of specificity could be important when a lectin crosslinks glycoproteins on cell surfaces, which often initiates signal transduction cascades.[4] Other lectin families (the galectins and spermadhesins) display quaternary structure variations that are strikingly similar to those of the legume lectins.[5] In the case of the daffodil and snowdrop lectins, a difference in quaternary structure appears to be related directly to a difference in carbohydrate specificity.[6]

DB58 is the heterodimeric lectin from the vegetative tissues of the tropical legume *Dolichos biflorus*.[7,8] It is closely related to DBL, the heterotetrameric 110 kDa seed lectin from the same plant.[9] Both proteins exhibit subunit heterogeneity that stems from post-translational removal of up to 12 amino acid residues from the C terminus of the newly synthesised 253 amino acid residue chains.[10] DBL forms a tetramer similar to that observed for PHA-L (the leucoagglutinating lectin from the common bean),[11] SBA (the agglutinin from soybean),[12] *Vicia villosa* isolectin B4[13] and isolectin II from *Ulex europaeus* (UEA-II).[14] The DBL-type dimer can be split conceptually into two types of dimer.[15] One of these is the canonical dimer, which has been observed in the lectins from pea and lentil, and is part of the ConA tetramer. The other, non-canonical dimer had not been observed separately before.

We present a high-resolution (2.5 Å) crystal structure of uncomplexed DB58, along with improved structures of PHA-L and SBA, and compare these proteins with other lectins with related quaternary structures. We investigate the interactions in the canonical dimer interface and the different ways it can be disrupted in the lectins with a different arrangement of the monomers.

Results and Discussion

Overall structure of DB58

The asymmetric unit of the DB58 crystals contains six monomers (A through F). Because of the non-crystallographic symmetry (NCS) restrictions that were applied during refinement, they are almost identical (but see below for an important exception). Four of the DB58 monomers (A, B, C and D) are associated in a quaternary arrangement like that of DBL (Figure 1). The presence of a separate non-canonical dimer formed by the remaining monomers (E and F) in this crystal proves that this dimer can exist independently and must be the quaternary structure of the protein in solution. Some legume lectins possess a hydrophobic binding site with high affinity for adenine and certain adenine-derived plant hormones.[16,17] This site and the monosaccharide-binding site are distinct and do not interact in any way. Binding of adenine has been described mainly for tetrameric legume lectins, including DBL, PHA-E and SBA. However, DB58, although dimeric, also binds adenine.[18] The adenine-binding site is located in the non-canonical interface of the DBL-type tetramer.[15] The fact that DB58 forms a non-canonical dimer is therefore strongly supported by the observation of adenine-binding by this protein.

DB58 has two potential glycosylation sites (Asn12-Ser-Ser and Asn79-Lys-Ser), which are both found to be glycosylated in biochemical tests.[7] There is no interpretable density for the Asn12 glycan. The central GlcNAc residue of the Asn79 glycan was visible in monomers B and E, and the entire core trisaccharide could be observed clearly in monomer F. Monomers A, D and F did not show any density for this glycan, which is close to the monosaccharide-binding site. The metal-binding sites of DB58 were identical with those of the other legume lectins, as expected. The strongly conserved *cis* peptide bond is present between Ala84 and Asp85. The four water ligands of the metals were introduced into the model, using a consensus geometry for the coordination bonds based on available high-resolution structures of other legume lectins.

The non-canonical dimer interface

The two monomers in a non-canonical dimer form two interfaces: a ββ interface between two equivalent β strands and an αβ interface between an α helix and part of a β sheet (Figure 1).[15] The ββ interface of DB58 is almost identical with that in DBL. It is formed by the interdigitation of the side-chains of residues 187, 189 and 191 in two equivalent β strands of the opposing monomers. The αβ interface is formed between the two-turn C-terminal α helix (residues 242-250) of a non-truncated monomer and part of the back sheet (β strands 2-8, 66-71 and 223-233) of the other, truncated monomer. In the DBL structure, the helices, which fit in a central cavity of the tetramer, could be associated unambiguously with specific subunits, but in the low-resolution DB58 structure they could not be modelled.

There is room for only one helix in the interface, but this helix can be contributed by either of the

† http://www.cermav.cnrs.fr/databank/lectine/

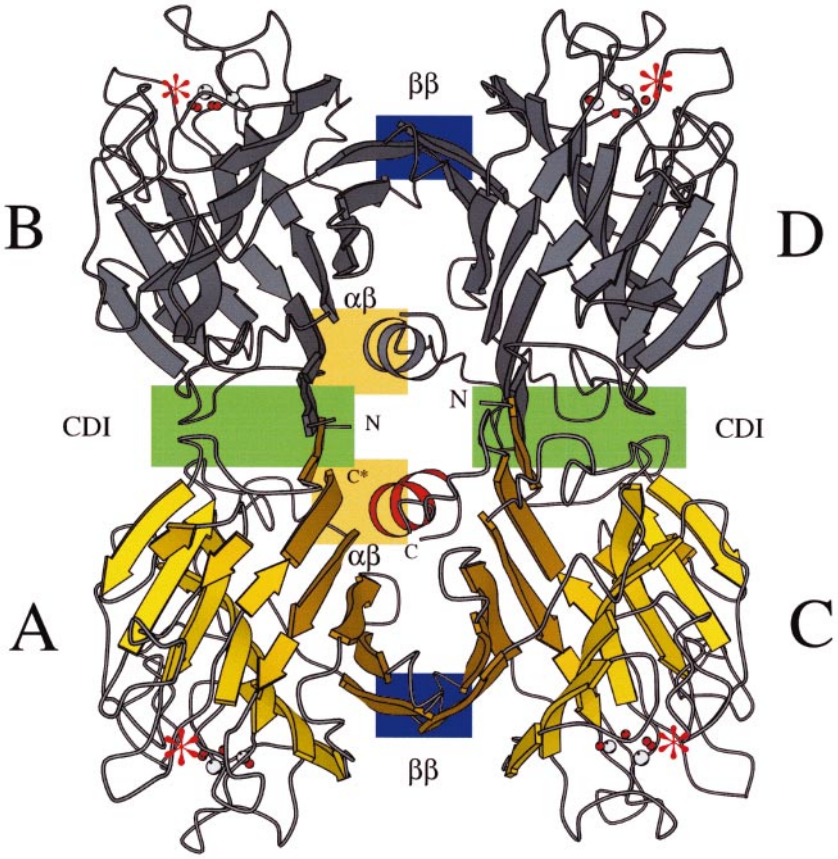

Figure 1. Overall structure of the DB58 tetramer (ABCD) observed in the crystal. This tetramer is very similar to the DBL tetramer and contains both canonical (AB and CD) and non-canonical (AC in colours and BD in grey) dimers. The asymmetric unit of the DB58 crystals contains an additional separate non-canonical EF dimer. The different subunit interfaces are labeled (**CDI** for the canonical dimer interface; αβ and ββ for the two parts of the non-canonical dimer interface). The back sheets are in orange, the front sheets are in yellow. The α helix in the non-canonical dimer interface is in red. The metal ions (grey spheres) and their water ligands (red spheres) are next to the sugar-binding sites, which are marked by asterisks. The N termini (**N**), the normal C termini (**C**) and the truncated C termini (**C***) are labeled.

two monomers, resulting in two possible, antiparallel orientations for the helix in the crystal. The helix is approximately symmetrical and can fit equally well into the grooves of the identical β sheets on either side (Figure 2). The density in the cavity was not very clear, and could not be accounted for by either orientation of the helix in the crystal. It was therefore considered that there might be statistical disorder in the crystal, and the helix was introduced in both orientations with half-occupancy. The calculated density thus ends up as the average of the contributions of the two orientations. The introduction of the helices reduces both the crystallographic *R*-factor and the free *R*-factor by approximately 1.2 %. The clearest density corresponds well to those regions where both orientations contribute atoms. For instance, the side-chain of Tyr247 from one orientation overlaps with the side-chain and main chain of Ala245 from the other orientation. Similar pairs are Leu242/Val249, Asp243/Asn251 and Ile244/Leu248.

The monosaccharide-binding sites of DB58

Figure 3 compares the monosaccharide-binding sites of DB58 and DBL. Three of the loops that contribute to the site, including the metal-binding loop and the *cis* peptide region, have identical conformations in these proteins. There are, however, differences in the side-chains at positions 102 and 127. Ser102 and Phe127 in DB58 correspond to Gly102 and Leu127 in DBL, respectively, which also means DB58 has the strongly conserved large aromatic side-chain that is missing from DBL.[15] The sequence of the specificity loop proper is identical in these proteins. However, because of the presence of Leu42 in an adjacent loop in DB58, the loop is shifted over a distance of approximately

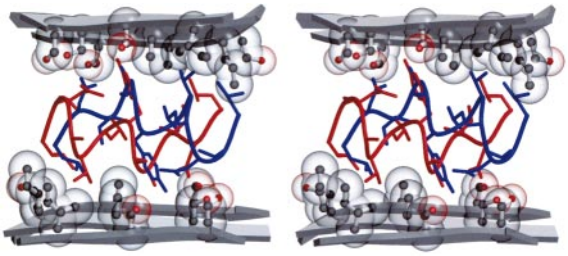

Figure 2. Stereo view of the two antiparallel orientations (in red and blue) of the central α helix in the DB58 crystal. Also shown are the surrounding strands from the back sheets sandwiching the helix and the atoms on the near faces of these sheets. The helix is sufficiently symmetric to fit into the grooves of the two identical sheets.

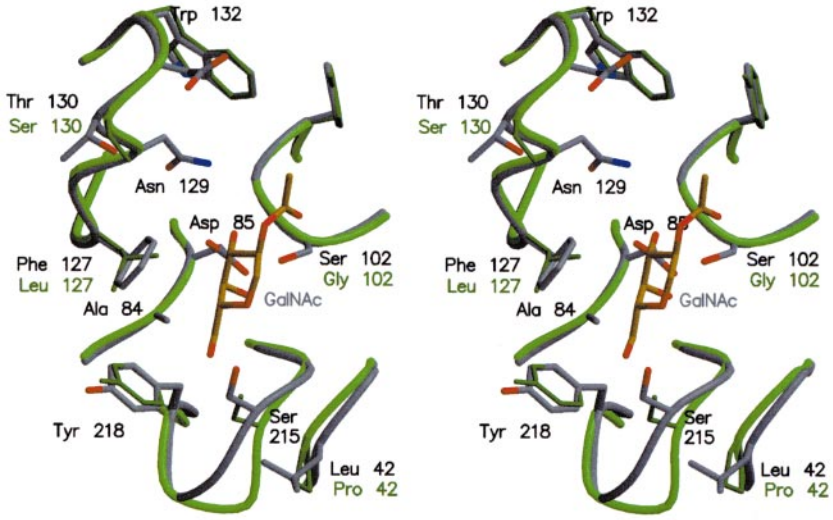

Figure 3. Stereo representation of a superposition of the monosaccharide-binding sites of DB58 (blue) and DBL (green). A GalNAc residue is shown as it is bound in the complex between DBL and the Forssmann disaccharide. The Figure shows how all the loops contributing to the site have identical conformations, except for the specificity loop (212-220), which is affected by the replacement of Pro42 in DBL by Leu42 in DB58. Other differences involve the side-chains of Ser102 and Phe127 in DB58, which correspond to Gly102 and Leu127 in DBL.

1 Å when compared to DBL, which has a more compact proline side-chain at the position corresponding to Leu42 in DB58. The O^{γ} atom of Ser215 is moved over the greatest distance (1.5 Å) and would clearly clash with a galactose residue bound in the same position and orientation as in DBL. These observations cannot be interpreted further without additional information about the specificity pattern of DB58 or structures of complexes between DB58 and suitable carbohydrates.

DBL is known to agglutinate red blood cells of blood group A and to precipitate derived glycoconjugates, whereas DB58 does not display these activities.[9] The distance between the two monosaccharide-binding sites of a DB58 dimer is approximately 55 Å and both sites are located on the same side of the dimer. The corresponding distance in a canonical dimer is approximately 68 Å. Additionally, there are differences in the orientation of the binding sites, as illustrated in Figure 4. By constructing a vector between the two residues of the Forssman disaccharide, as bound by DBL, the relative orientation of the binding sites in a DBL-type tetramer was quantified. The angle between these vectors is clearly obtuse for the canonical dimer, and clearly acute for the DB58-type dimer. A DBL-type tetramer combines both these geometries and has a distance of about 86 Å between the diagonally opposed monomers. For the GS-IV and EcorL dimers, the binding sites point away from each other and the ''exit vectors'' lie almost on a straight line. Among the four different dimers, the binding sites of the DB58 type have the smallest angle and the smallest distance between them. These differences in geometry might influence the ability of the DB58 dimer to bridge large distances and crosslink multivalent carbohydrate ligands. It could bind ligands close together on a surface and cover it, but would have a harder time crosslinking ligands on different cells than the other three types of dimers. The geometry and freedom of movement of the epitopes on the carbohydrate ligand would, of course, also affect these interactions.

Overall structure of PHA-L and SBA

Some minor improvements were introduced into the existing PHA-L model[11] based on the expanded diffraction data. The improved electron density is due to the introduction of additional diffraction data and recent progress in refinement and calculation protocols, such as bulk solvent and global anisotropic *B*-factor corrections. A comparison of the central cavities of the PHA-L tetramer and the DB58 crystal tetramer (data not shown) clearly indicates that the electron density for PHA-L does not warrant the inclusion of α helices. The αβ interface appears not to be essential for the stability of the PHA-L tetramer.

The asymmetric unit of the SBA crystals contains only one monomer and one pentasaccharide.[12] The SBA tetramers and the larger lattice of pentasaccharide-bridged tetramers are formed by the crystal symmetry operations. The observed electron density is averaged over all four monomers, and conformational differences for loops or helices cannot be determined. The previously uninterpreted density in the central cavity corresponds well to the average of two antiparallel orientation of an α helix as in DB58. The α helix and the pentasaccharides were both introduced into the model in one of the two possible orientations, with half-occupancy for all the atoms. Application of the symmetry operations yielded the alternative orientations of these two elements, averaging both orientations, and closely modelling the statistical disorder in the crystal. In this case too, clearer electron density made possible a number of small improvements to the model, including the addition of several side-chains and the central galactose residue of the ligand Gal-β1,4-GalNAc-β1,2-[Gal-β1,4-

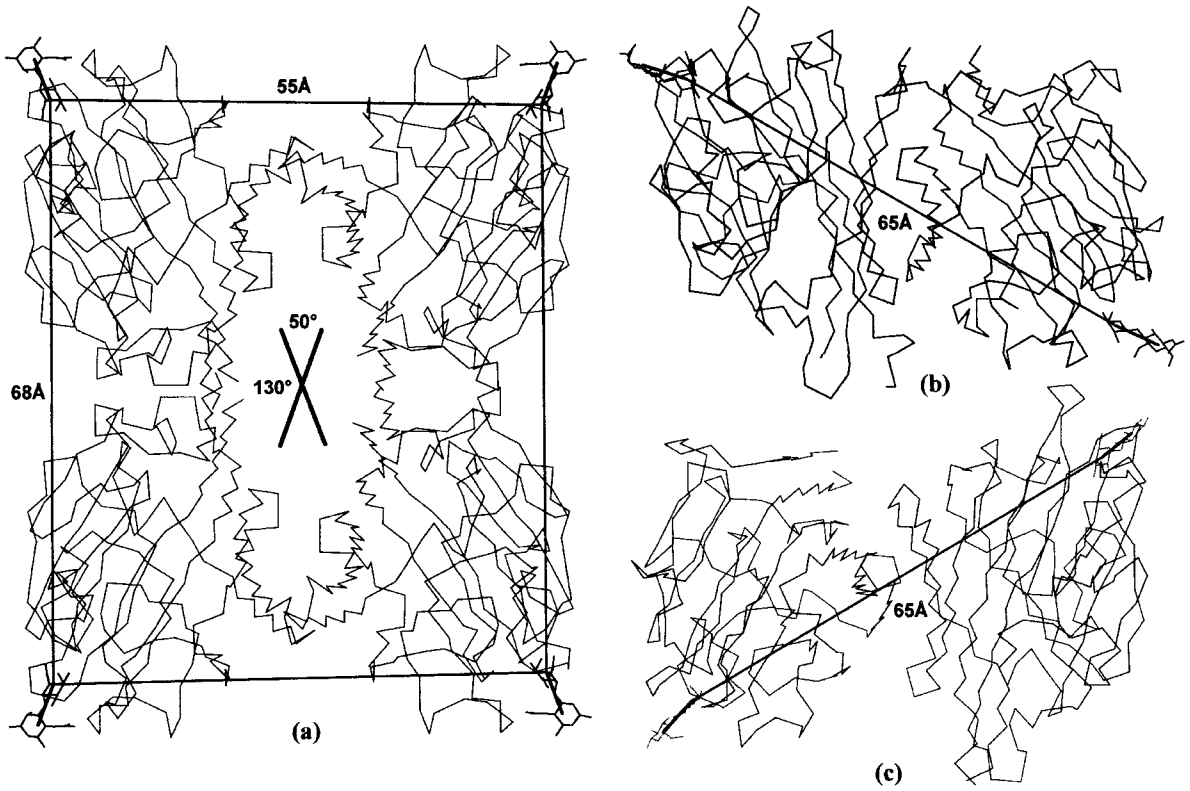

Figure 4. Comparison of the relative positions and orientations of the monosaccharide-binding sites in the four types of legume lectin dimer. For each dimer type, a complex with an oligosaccharide was chosen: (a) the Forssmann disaccharide for DBL, (b) the Le^b antigen for GS-IV and (c) lactose for EcorL. A reference vector was constructed from the geometrical midpoint of the residue bound in the monosaccharide-binding site to the midpoint of the next residue, in order to describe the orientation of the disaccharide. The angle between the vectors in each dimer type was calculated. Although the precise values of the angles depend on the choice of reference points for the vectors, it is clear that the angle is acute (around 50°) for the DB58-type dimer, obtuse for the canonical dimer (around 130°) and around 180° for the GS-IV and EcorL dimers. The distance between the binding sites is similar (around 65 Å) for the EcorL and GS-IV dimers.

GalNAc-β1,3-]Gal, which is now completely visible.

Comparison of the canonical dimer contacts

The association of two DB58-type dimers into a DBL-type tetramer is equivalent to the formation of two canonical dimers and involves the antiparallel association of two β strands, as well as three different loop contacts (*ab*, *cc* and *ba*; Figure 5). These contacts were compared in DBL and DB58, and found to be essentially identical, except for the contact between loops *a* and *b*. As refinement of DB58 progressed, it became clear that loop *a* (Phe11-Phe16) exhibits conformational flexibility. NCS restrictions for this region were released, and significant differences between the monomers could be modelled (see Figure 5 for a superposition of the different conformations of loop *a*). In monomers A, B, C and D, loop *a* is involved in the canonical dimer contacts, whereas in monomers E and F it is free in the crystal solvent. DBL has a proline residue at position 14, whereas DB58 has Ser14 at the corresponding position. Loop *a* in DBL has a well-defined conformation and the hydrophobic Pro14 side-chain is packed against the aromatic side-chain of Tyr203 in loop *b* of the opposing monomer. In contrast, the serine residue of DB58 does not offer a significant hydrophobic surface to pack against Tyr203. The other contacts in the canonical interface are identical between DBL and DB58, and it would appear that the difference of one amino acid residue at position 14 is a subtle but sufficient mechanism to cause DB58 to be dimeric and DBL to be tetrameric in solution. Another case where a single amino acid substitution has been found to affect oligomeric association is the His47Asn mutant of triosephosphate isomerase (H47N-TIM).[19]

The presence of the tetramer in the crystal indicates that association of DB58 dimers is not sterically prohibited and can occur in suitable conditions. For both DBL and DB58 there is presumably a dynamic equilibrium between dimers and tetramers in solution, with a strong preference for tetramers in the case of DBL and for dimers with DB58. There are numerous other examples of higher-order multimer formation in lectin crystal

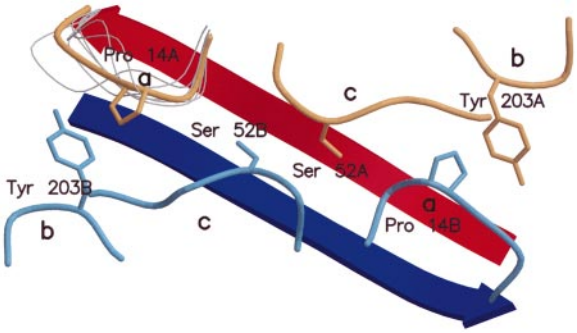

Figure 5. The different contacts in the canonical dimer interface. The image is based on the structure of DBL and the three relevant loops are labelled *a* (11-16 in DBL), *b* (193-203) and *c* (50-66). The only difference between DBL and DB58 in these regions is the replacement of Pro14 of DBL by Ser14 in DB58. The *a* loop in DB58 is flexible, and a number of different backbone conformations are indicated in thin grey lines.

structures (summarised by Hamelryck *et al.*)[20] and in solution.[21,22] A particularly striking example is FRIL (from *Dolichos lablab*), which is a dimeric lectin in solution and can form a weakly associated ConA-type tetramer in a cross-linked complex with a sugar.[20] In this case, the structural basis of the marginal stability of the FRIL tetramer (compared to the stable ConA tetramer) could not be pinpointed as clearly as in the DBL/DB58 case. The formation of the type II[3] crosslinked lattice between FRIL and a trisaccharide, observed in the crystal, depends on the tetrameric association of the protein. In DB58 too, a tetramer would clearly have greater crosslinking potential than the individual monomers, as described above. The ability of lectins like DB58 and FRIL to form tetramers could therefore be physiologically relevant in the context of receptor crosslinking, which often occurs in signal transduction systems.[4] This could allow for subtle regulation, based on the local concentration of the lectin on the cell surface, the local pH and other factors.

The change from proline to serine also introduces a glycosylation site in loop *a* of DB58. PHA-L is also glycosylated at the corresponding position, yet it does form the DBL-type tetramer. It has been established that the relative importance of glycosylation and factors intrinsic to the protein in determining quaternary structure can vary significantly.[23,24] In order to rationalise the quaternary structure observations, Prabu *et al.*[25] have calculated global parameters (hydrophobic surface buried on association, surface complementarity and model-based interaction energy) for the natural legume lectin oligomers and a number of artificially constructed associations. The DBL/DB58 case illustrates that very subtle, local differences can be enough to destabilise an interaction. For this reason, we attempted a number of detailed structure analyses of other lectins that do not form canonical dimers.

Figure 6 shows the sequence alignment for the relevant sections. Important differences occur in loop *c*, which interacts with its counterpart in the opposing monomer. The small neutral side-chain (alanine or serine) in the canonical dimer lectins is replaced by a larger, positively charged lysine residue in EcorL (the seed lectin from *Erythrina corallodendron*)[23] and WBA-I (the first basic agglutinin from *Psophocarpus tetragonolobus*),[24] and by a negatively charged glutamate residue in isolectin IV of *Griffonia simplicifolia* (as observed by Delbaere *et al.*).[26] These charged residues would be buried in the predominantly hydrophobic canonical dimer interface. In the *Griffonia* lectin (GS-IV), there is an

	loop a	loop b	loop c	
(DB58 numbering)	11 16	203	52	canonical dimer
DB58	FSFKNFN--S**S**SFILQG	VDITNELPE**Y**VSIGFSATT	GRAFYS**S**PIQIYD-SSA	*only in crystal*
DBL	FSFKNFN--S**P**SFILQG	VDITNELPE**Y**VSIGFSATT	GRAFYS**S**PIQIYD-SSA	*stable*
PHA-L	FNFQRFN--ETNLILQR	VDITNELPE**Y**VSVGFSATT	GRAFYS**S**PIQIYDKSTG	*stable*
PHA-E	FSFQRFN--ETNLILQR	VDLKSVLPE**W**VSVGFSATT	GRAFYS**A**PIQIWDNTTG	*stable*
SBA	FSWNKFVPKQPNMILQG	VDLKSVLPE**W**VIVGFTATT	GRAFYS**A**PIQIWDNTTG	*stable*
UEA-II	FNFDKFVPNQKNIIFQG	VDLKAILPE**W**VSVGFSGGV	GRALYA**A**PIQIWD-SAT	*stable*
MAL	FTINNFVPNEADLLFQG	VDLKEILPE**W**VRVGFSAAT	GRALYA**A**PVRIWGNTTG	*unknown*
UEA-I	FKFKNFSQNGKDLSFQG	VDLKEILPE**W**VSVGFSG--	GIARYA**A**PIHIWNCNTG	*unknown*
GS-IV	YSLKNGT--EITFL--G	VDLAKVLPQ**K**VRIGFSA--	GQASYS**E**PVFLWD----	*not formed*
EcorL	FSFSEFEPGNDNLTLQG	VDVKQVLPE**W**VDVGLSGAT	GRTLYA**K**PVHIWDMTTG	*not formed*
WBA-I	FNFNQFHQNEEQLKLQR	VDLKQVLPE**S**VNVGFSAAT	GRALYA**K**PVHIWDMTTG	*not formed*

Figure 6. Structure-based sequence alignment for legume lectins. Alignments were performed with BLASTP 2.0.5 and locally verified by comparing the crystal structures. *N*-glycosylation sites (NX[S/T]) have been underlined. Key residues are marked in bold.

additional replacement of the highly conserved aromatic residue in the middle of the *b* loop by a lysine residue, which would clash with the *a* loop of the opposing monomer in a canonical dimer. In WBA-I, this aromatic side-chain is replaced by a much smaller serine side-chain.

Like DB58, the leucoagglutinating lectin from *Maackia amurensis* (MAL) is known to be dimeric in solution, but it forms DBL-type tetramers in the crystals used for the structure determination.[27] The MAL crystals do not contain isolated dimers, so it is still not known whether the lectin forms canonical or DB58-type dimers in solution or which type of dimer is formed only in the crystal. For this *Maackia* lectin and isolectin I from *Ulex europaeus* (UEA-I), given as additional examples in Figure 6, there is no obvious sequence element that would prohibit canonical dimer formation. If the geometry of the DB58-type dimer indeed prohibits agglutination, the observation that MAL has agglutination activity would mean that it would have to form a canonical dimer in solution, which in turn would imply that it is the DB58-type dimer that is formed in the crystal.

The stable quaternary structure of a lectin oligomer is determined ultimately by the relative stability of the different interfaces. There are different ways to affect the stability of these interfaces. The canonical dimer interface can be destabilised by sterical hindrance or by a conformationally flexible loop. The non-canonical interface can be stabilised in different ways, for instance by an α helix (as in DBL/DB58 and SBA) or by a disulphide bond (as in UEA-II).[14] It is still difficult to predict the quaternary structure of a lectin based on the sequence, especially in very subtle cases like the DBL/DB58 or FRIL systems.

Materials and Methods

Crystallization, data collection and processing

Purification and crystallization of uncomplexed natural DB58 have been described elsewhere.[28] Briefly, suitable crystals were obtained using the ''hanging drop'' method with 5-7 mg ml^{-1} protein, 100 mM sodium citrate buffer (pH 5.6), 8-12 % (w/v) PEG6000 and 0.2-0.4 M sodium chloride. Crystallisation of uncomplexed PHA-L and the SBA/pentasaccharide complex have been described elsewhere.[12,29] The properties of the different crystals and the data collection statistics are summarized in Table 1. Data for the DB58 crystals were collected on station BW7A of the DESY synchrotron in Hamburg, Germany ($\lambda = 0.98$ Å). Diffraction images were processed using the programs Denzo, XDisplayF and ScalePack from the HKL Package[30] and the CCP4 program Truncate.[31] The structure factors for the SBA complex[12] were retrieved from the Protein Data Bank (access code r1SBFsf). The CCP4 programs Scala, Truncate and MTZ2various were used to reprocess these datasets. Denzo, ScalePack, Scala and Truncate were used to process previously unused diffraction images for PHA-L and combine them with the structure factors that were already available. The completeness for the existing PHA-L dataset was 88.5 %,[11] whereas the new combined data set reaches 94.4 %.

Model building and refinement

Refinement was done using the simulated annealing torsion angle refinement and conjugate gradient least-squares refinement protocols of the CNS program[32] with the maximum likelihood target function (MLF target) based on structure factors. Throughout the refinement, real space bulk solvent and anisotropic *B*-factor corrections were used. For DB58, NCS restraints ($\sigma_b = 0.5$ $Å^2$; weight 1000 kcal mol^{-1} $Å^2$) were applied between the six monomers in the asymmetric unit. In the later stages of refinement, NCS restraints were released for certain regions of the subunits. This did not lead to different

Table 1. Properties of the different crystals and data collection statistics for native DB58, native PHA-L and the SBA/pentasaccharide complex

	DB58	PHA-L	SBA
Space group	$P2_12_12_1$	$C2$	$I4_122$
Cell parameters			
a (Å)	100.81	106.48	122.64
b (Å)	130.28	121.28	122.64
c (Å)	137.06	89.96	90.56
β (deg.)	90.0	93.7	90.0
Contents of the asymmetric unit	Three dimers	One tetramer	One monomer and one pentasaccharide
Resolution			
Range (Å)	20-2.5	20-2.8	20-2.2
Final shell (Å)	2.59-2.5	2.95-2.8	2.25-2.2
$\langle I/\sigma_I \rangle$ (total)	12.9	11.23	12
$\langle I/\sigma_I \rangle$ (final shell)	2.78	3.2	3.7
R_{merge} (%)	10.4	10.7[a]	12.8[b]
Observed reflections	259,246	100,006	57,894
Unique reflections	59,075	26,546	19,716
Global completeness (%)	93.8	94.4	80.0
Completeness in the final shell (%)	95.7	82.1	72.7
Redundancy	4.4	3.8	2.9

[a] Global value for the combined data.
[b] From Olsen *et al.*[12]

Table 2. Quality parameters of the final models for DB58, PHA-L and SBA

	DB58	PHA-L	SBA
R_{cryst} (%)	22.1	18.2	19.0
R_{free} (%)	24.7	21.2	23.0
r.m.s.d. from ideal Bond lengths (Å)	0.009	0.008	0.007
r.m.s.d. from ideal Angles (°)	1.661	1.555	1.351
r.m.s.d. from ideal Dihedral angles (deg.)	26.774	27.151	26.056
r.m.s.d. from ideal Improper angles (deg.)	0.871	0.847	0.796
Distribution of the residues in the Ramachandran plot (%)			
Core	84.1	85.2	85.1
Allowed	14.4	14.2	13.5
Additional allowed	1.3	0.5	1.0
Forbidden	0.2	0.2	0.5

conformations in the different monomers, except in the case of the loop between Phe11 and Phe16, where different conformations could be modelled in the different monomers. NCS restrictions were also applied to the four monomers in the asymmetric unit of the PHA-L crystal.

Visualisation and model building were done with the program TURBO-FRODO starting from the existing low-resolution DB58 model[15] and the deposited models for PHA-L and SBA. Figures were made with the programs MOLSCRIPT,[33] BobScript[34] and Raster3D.[35] The quality and secondary structure of the models were analysed with the CCP4 programs PROCHECK[36] and PROMOTIF.[37] The quality parameters of the final models are summarized in Table 2. NAccess was used to calculate accessible surfaces and determine buried hydrophobic surfaces. Models were compared by superimposing them using a least-squares algorithm with the CCP4 program LSQKab. The following PDB entries were used for structure comparisons: 1DBN (MAL), 1AX0 (EcorL), 1GSL (GS-IV), 1LU1 and 1LU2 (DBL), 1QMO (FRIL), 1WBL (WBA-I) and 1QNW (UEA-II).

Protein Data Bank accession numbers

Coordinates and structure factors have been submitted to the RCSB Protein Data Bank with accession codes 1G7Y and r1G7Ysf for uncomplexed DB58, 1G8W and r1G8Wsf for uncomplexed PHA-L and 1G9F and r1G9Fsf for SBA.

Acknowledgements

L.B. is a research assistant with the Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (F.W.O.). R.L. and T.H. are postdoctoral fellows of the F.W.O. This work was supported by the Vlaams Interuniversitair Instituut voor Biotechnologie (V.I.B.). We thank Maria Vanderveken and Yves Geunes for excellent technical assistance. The EMBL provided essential beamtime on the DORIS storage ring at the DESY synchrotron in Hamburg, Germany.

References

1. Lis, H. & Sharon, N. (1998). Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637-674.
2. Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). Legume lectin structure. *Biochim. Biophys. Acta,* **1383**, 9-36.
3. Brewer, C. F. (1996). Multivalent lectin-carbohydrate cross-linking interactions. *Chemtracts Biochem. Mol. Biol.* **6**, 165-179.
4. Heldin, C. H. (1995). Dimerization of cell surface receptors in signal transduction. *Cell,* **80**, 213-223.
5. Bouckaert, J., Hamelryck, T., Wyns, L. & Loris, R. (1999). Novel structures of plant lectins and their complexes with carbohydrates. *Curr. Opin. Struct. Biol.* **9**, 572-577.
6. Chandra, N. R., Ramachandraiah, G., Bachhawat, K., Dam, T. K., Surolia, A. & Vijayan, M. (1999). Crystal structure of a dimeric mannose-specific agglutinin from garlic: quaternary association and carbohydrate specificity. *J. Mol. Biol.* **285**, 1157-1168.
7. Etzler, M. E. (1994). Isolation and characterisation of subunits of DB58, a lectin from the stems and leaves of *Dolichos biflorus*. *Biochemistry*, **33**, 9778-9783.
8. Schnell, D. J. & Etzler, M. E. (1988). cDNA cloning, primary structure, and *in vitro* biosynthesis of the DB58 lectin from *Dolichos biflorus*. *J. Biol. Chem.* **263**, 14648-14653.
9. Etzler, M. E. (1996). The *Dolichos biflorus* lectin family: a model system for studying legume lectin structure and function. In *Lectins-Biology, Biochemistry, Clinical Biochemistry* (Van Driessche, E., Rougé, P., Beeckmans, S. & Bøg-Hansen, T. C., eds), vol. 11, pp. 3-9, Textop, Denmark.
10. Young, N. M., Watson, D. C., Yaguchi, M., Adar, R., Arango, R., Rodriguez-Arango, E., Sharon, N., Blay, P. K. S. & Thibault, P. (1994). C-terminal post-translational proteolysis of plant lectins and their recombinant forms expressed in *Escherichia coli*. *J. Biol. Chem.* **270**, 2563-2570.
11. Hamelryck, T. W., Dao-Thi, M., Poortmans, F., Chrispeels, M. J., Wyns, L. & Loris, R. (1996). The crystallographic structure of phytohemagglutinin-L. *J. Biol. Chem.* **271**, 20479-20485.
12. Olsen, L. R., Dessen, A., Gupta, D., Sabesan, S., Sacchettini, J. C. & Brewer, C. F. (1997). X-ray crystallographic studies of unique cross-linked lattices

between four isomeric biantennary oligosaccharides and soybean agglutinin. *Biochemistry,* **36**, 15073-15080.
13. Osinaga, E., Tello, D., Batthyany, C., Bianchet, M., Tavares, G., Duran, R., Cervenansky, C., Camoin, L., Roseto, A. & Alzari, P. M. (1997). Amino acid sequence and three-dimensional structure of the Tn-specific isolectin B4 from *Vicia villosa*. *FEBS Letters,* **412**, 190-196.
14. Loris, R., De Greve, H., Dao-Thi, M.-H., Messens, J., Imberty, A. & Wyns, L. (2000). Structural basis of carbohydrate recognition by lectin II from *Ulex europaeus*, a protein with a promiscuous carbohydrate binding site. *J. Mol. Biol.* **301**, 987-1002.
15. Hamelryck, T. W., Loris, R., Bouckaert, J., Dao Thi, M.-H., Strecker, G., Imberty, A., Fernandez, E., Wyns, L. & Etzler, M. E. (1999). Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. *J. Mol. Biol.* **286**, 1161-1177.
16. Roberts, D. D. & Goldstein, I. J. (1983). Adenine binding sites of the lectin from Lima beans (*Phaseolus lunatus*). *J. Biol. Chem.* **258**, 13820-13824.
17. Maliarik, M. J. & Goldstein, I. J. (1988). Photoaffinity labeling of the adenine binding site of the lectins from Lima bean, *Phaseolus lunatus* and the kidney bean, *Phaseolus vulgaris*. *J. Biol. Chem.* **263**, 11274-11279.
18. Gegg, C. V. & Etzler, M. E. (1994). Photoaffinity labelling of the adenine binding sites of two *Dolichos biflorus* lectins. *J. Biol. Chem.* **269**, 5687-5692.
19. Borchert, T. V., Zeelen, J. P., Schliebs, W., Callens, M., Minke, W., Jaenicke, R. & Wierenga, R. K. (1997). An interface point-mutation variant of triosephosphate isomerase is compactly folded and monomeric at low protein concentrations. *FEBS Letters,* **367**, 315-318.
20. Hamelryck, T. W., Moore, J. G., Chrispeels, M. J., Loris, R. & Wyns, L. (2000). The role of weak protein-protein interactions in multivalent lectin-carbohydrate binding: crystal structure of cross-linked FRIL. *J. Mol. Biol.* **299**, 875-883.
21. Cho, M. & Cummings, R. D. (1995). Galectin-1, a β-galactoside-binding lectin in Chinese hamster ovary cells. *J. Biol. Chem.* **270**, 5198-5206.
22. Calvete, J. J., Thole, H. H., Raida, M., Urbanke, C., Romero, A., Grangeiro, T. B., Ramos, M. V., Almeida de Rocha, I. M., Guimarãs, F. N. & Cavada, B. S. (1999). Molecular characterisation and crystallisation of *Diocleinae* lectins. *Biochim. Biophys. Acta,* **1430**, 367-375.
23. Shaanan, B., Lis, H. & Sharon, N. (1991). Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science,* **254**, 862-866.
24. Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M. & Suguna, K. (1998). Carbohydrate specificity and quaternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 Å resolution. *J. Mol. Biol.* **276**, 787-796.
25. Prabu, M. M., Suguna, K. & Vijayan, M. (1999). Variability in quaternary associations of proteins with the same tertiary fold: a case study and rationalization involving legume lectins. *Proteins: Struct. Funct. Genet.* **35**, 58-69.
26. Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Wilson, K. S. & Dauter, Z. (1993). Structures of the lectin IV of *Griffonia simplicifolia* and its complex with the Lewis b human blood group determinant at 2.0 Å resolution. *J. Mol. Biol.* **230**, 950-965.
27. Imberty, A., Gauthier, C., Lescar, J., Perez, S., Wyns, L. & Loris, R. (2000). An unusual carbohydrate binding site revealed by the structures of two *Maackia amurensis* lectins complexed with sialic acid containing oligoscaccharides. *J. Biol. Chem.* **275**, 17541-17548.
28. Dao-Thi, M., Hamelryck, T. W., Bouckaert, J., Körber, F., Burkow, V., Poortmans, F., Etzler, M., Strecker, G., Wyns, L. & Loris, R. (1998). Crystallisation of two related lectins from the legume plant *Dolichos biflorus*. *Acta Crystallog. sect. D,* **54**, 1446-1449.
29. Dao-Thi, M., Hamelryck, T. W., Poortmans, F., Voelker, T. A., Chrispeels, M. J. & Wyns, L. (1996). Crystallisation of glycosylated and nonglycosylated phytohemagglutinin-L. *Proteins: Struct. Funct. Genet.* **24**, 134-137.
30. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
31. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.
32. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D,* **54**, 905-921.
33. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
34. Esnouf, R. M. (1997). An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132-134.
35. Merrit, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallog. sect. D,* **50**, 869-873.
36. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
37. Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF - a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212-220.

Edited by R. Huber

(Received 5 December 2000; received in revised form 21 March 2001; accepted 21 March 2001)