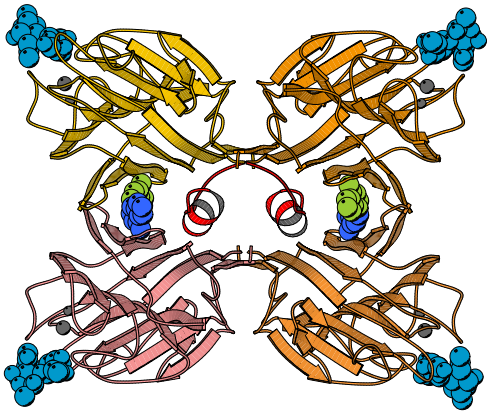
Structural studies of four members of the legume lectin family: Phytohemagglutinin-L, Arcelin-5, *Dolichos biflorus* seed lectin and *Dolichos biflorus* stem and leaf lectin



Thomas W. Hamelryck
Laboratorium voor Ultrastructuur
Vrije Universiteit Brusssel (VUB)

Defended on March 5th, 1999

Supervisor: Prof. Lode Wyns

Evaluating committee
Prof. Kurt Drickamer, University of Oxford
Prof. Jan Kroon, Utrecht University
Prof. Raymond Cunin, VUB
Prof. Rudy Willem, VUB

Contents

1 Introduction **3**
- 1.1 What are lectins? . . . 3
- 1.2 The legume lectins . . . 3
- 1.3 The legume lectin monomer . . . 4
- 1.4 Sugar binding by legume lectins . . . 5
- 1.5 Quaternary structure . . . 9
- 1.6 Hydrophobic binding sites in legume lectins . . . 12
- 1.7 Truncated legume lectins . . . 12
- 1.8 Legume lectin structures: an overview . . . 13
- 1.9 Homologues outside the *Leguminosae* . . . 14
- 1.10 Conclusions . . . 15
- 1.11 References . . . 15

2 PHA-L **22**
- 2.1 Introduction . . . 22
- 2.2 Materials and methods . . . 23
  - 2.2.1 Crystallization and data collection . . . 23
  - 2.2.2 Structure determination and interpretation . . . 24
- 2.3 Results and discussion . . . 25
  - 2.3.1 Structure description . . . 25
  - 2.3.2 Quaternary structure . . . 26
  - 2.3.3 The metal and sugar binding sites . . . 28
  - 2.3.4 Comparison between PHA-L and SBA . . . 29
  - 2.3.5 Putative position of the adenine/cytokinin binding site 30
- 2.4 Conclusions . . . 31
- 2.5 References . . . 32

3 Arcelin-5 **36**
- 3.1 Introduction . . . 36
- 3.2 Materials and methods . . . 37
  - 3.2.1 Crystallization and data collection . . . 37
  - 3.2.2 Structure solution and refinement . . . 38
  - 3.2.3 Gel filtration . . . 40
- 3.3 Results and discussion . . . 40
  - 3.3.1 Overall structure . . . 40
  - 3.3.2 The quaternary structure of Arc5 . . . 41
  - 3.3.3 The glycan attached to Asn 22 . . . 42
  - 3.3.4 The truncated metal binding site . . . 44
- 3.4 Conclusions . . . 46
- 3.5 Notes . . . 47
- 3.6 References . . . 47

4 DBL and DB58 **50**
4.1 Introduction 50
4.2 Materials and methods 51
4.2.1 Preparation of the L127F seed lectin mutant 51
4.2.2 Binding assays 51
4.2.3 Crystallization 52
4.2.4 Data collection 52
4.2.5 Molecular replacement, model building and refinement 52
4.3 Results and discussion 55
4.3.1 Overall structure of DBL 55
4.3.2 Structure of the adenine binding site 56
4.3.3 DB58, a vegetative legume lectin from *D. biflorus* 59
4.3.4 Blood group A trisaccharide binding 60
4.3.5 GalNAc specificity 63
4.3.5.1 The role of Leu 127 63
4.3.5.2 The role of the specificity loop 65
4.3.6 Forssman disaccharide binding 66
4.3.7 A multi-purpose secondary binding site in the legume lectin family 67
4.4 Conclusions 68
4.5 References 69

5 Summary **73**

6 Samenvatting **75**

7 Abbreviations used **77**

Chapter 1

Introduction

Reference: Hamelryck, T.W., Loris, R., Bouckaert, J. & Wyns, L. (1998) Properties and structure of the legume lectin family. **Trends Glycosci. Glycobiol.**, 10, 349-404.

1.1 What are lectins?

Lectins are proteins of non-immune origin that bind carbohydrates in a reversible manner, without having any enzymatic activity towards this bound carbohydrate. The term lectin was first coined by Boyd & Shapleigh (1954) and is derived from the latin verb *legere*, which means "to choose". Lectins were originally discovered through their ability to agglutinate blood cells via cross linking of the sugar residues present on their surface. Lectins are ubiquitous in nature and can be found in viruses, bacteria, plants and animals, including humans. They are the natural partners of the various oligosaccharides that nature uses to "tag" cells via glycoproteins and glycolipids. Lectins play a role in important processes like infection, host defense, fertilization, cancer, protein transport and embryogenesis (Gabius & Gabius, 1997). They are also important as tools in biology, histology and medicine, e.g., as an aid in the purification of glycoproteins or the determination of blood groups (Van Damme *et al.*, 1998). Lectin-carbohydrate recognition is one of the three fundamental biological recognition mechanisms, the two others being protein-nucleic acid recognition and protein-protein recognition. A detailed understanding of lectin-carbohydrate recognition mechanisms is therefore of great practical as well as fundamental value.

1.2 The legume lectins

The legume lectins are a family of sugar binding proteins found in the seeds, and in smaller amounts in roots, stems, leaves and bark of plants belonging to the *Leguminosae* family (Sharon & Lis, 1990; Loris *et al.*, 1998). The exact function of the legume lectins *in vivo* has not yet been established although considerable attention has been devoted to the possibilities that they may be involved in the defense of plants against predators or in the interaction of the plant with *Rhizobium* symbionts (Chrispeels & Raikhel, 1991; Etzler, 1992; Brewin & Kardailsky, 1997). Recently, homologues in other plant families (Hervé *et al.*, 1996; Schröder *et al.*, 1995)

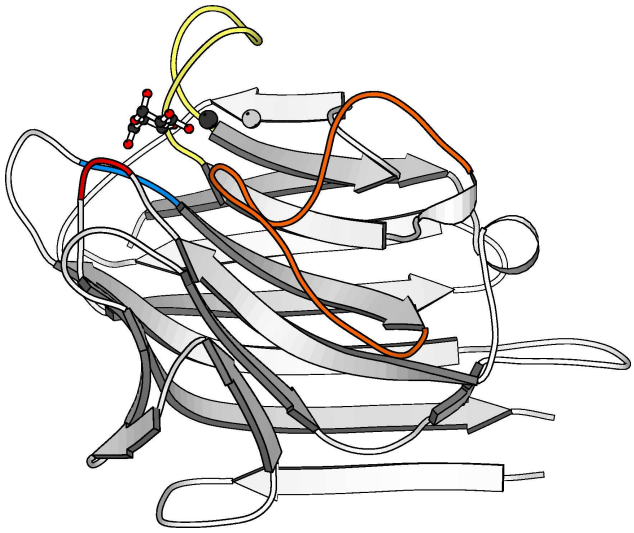


Figure 1.1: A typical legume lectin monomer (lentil lectin, PDB entry 1LES, Casset *et al.*, 1995). β-Strands are shown as arrows. The conserved monosaccharide binding site is located on the top of the monomer, in a shallow depression. The bound Glc molecule is shown as a ball-and-stick model. The four sugar binding loops are shown in different colors (Ala-Asp *cis*-peptide bond in blue; metal binding loop in yellow; Ω-loop in orange; monosaccharide specificity loop in red). The metals are shown as gray spheres (Ca^{2+} dark gray; Mn^{2+} light gray).

and in animals have been described (Fiedler & Simons, 1994; Itin *et al.*, 1996). They have been used for decades as a model system for the study of protein-carbohydrate interactions, because they show an amazing variety of binding specificities and are easy to obtain and purify. Over the years, a quite impressive amount of structural data has been gathered. For reviews with an emphasis on structural features, see (Rini, 1995; Weis & Drickamer, 1996; Loris *et al.*, 1998; Lis & Sharon, 1998).

1.3 The legume lectin monomer

Legume lectins are dimeric or tetrameric hetero- or homo-oligomers. The structures of the different legume lectin monomers are extremely similar, and their structure can be described as a β-sandwich consisting of a flat, six stranded sheet (the back sheet), and a curved, seven stranded sheet (the front sheet). A small third β-sheet, consisting of five strands, holds the two larger sheets together (Banerjee *et al.*, 1996). The sugar binding site is located on the concave side of the β-sandwich, formed by the curved front sheet, next to a double metal binding site (Fig. 1.1).

The topology of the monomer is related to the right handed class I jelly-roll fold (Stirk *et al.*, 1992), but it contains three insertions, one of which is directly involved in sugar and metal binding (Fig. 1.2). Essentially the same topology is found in the galectins (Lobsanov *et al.*, 1993; Liao *et al.*, 1994), the pentraxins (Emsley *et al.*, 1994) and the spermadhesins (Romao *et al.*, 1997; Varela *et al.*, 1997; Romero *et al.*, 1997), all of which are lectins. Many other sugar binding or processing proteins have a topology

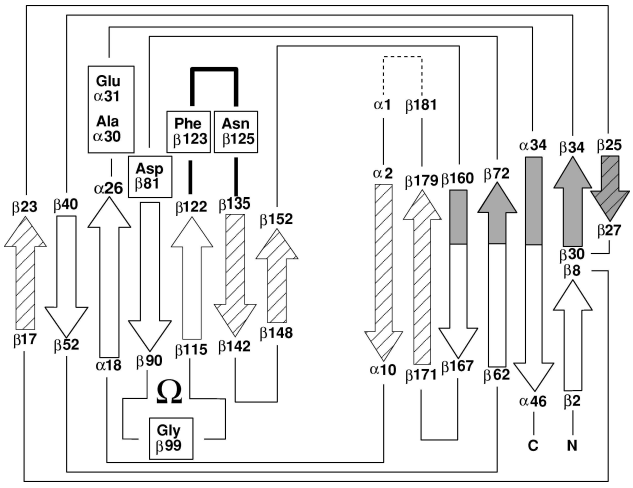


Figure 1.2: Topology diagram of a typical legume lectin monomer (lentil lectin, PDB entry 1LES). β-Strands are shown as arrows, connecting loops as simple lines. The monomer consists of a flat six strand β-sheet (shown right), a curved seven strand β-sheet (shown left) and a small connecting five strand β-sheet (shown in gray). The topology is related to the jellyroll fold, with three additional inserts (shown as hatched strands). One of these inserts contains two of the four conserved sugar binding residues (Phe 123β, Asn 125β) and the metal binding loop (shown in bold). The two other conserved residues are Asp 81β (involved in the Ala-Asp *cis*-peptide bond) and Gly 99β (present in a large β-loop that packs against the front sheet). Ala 30α and Glu 31α belong to the monosaccharide specificity loop. Lentil lectin belongs to the Glc specific *Vicieae* tribe lectins and is proteolitically cleaved at position 181 (shown as a dotted line), resulting in a two-chain lectin (chains α and β).

related to the jellyroll fold (Srinivasan *et al.*, 1996). These include e.g. β-D-glucanase (Keitel *et al.*, 1993), PNGase F (Norris *et al.*, 1994; Kuhn *et al.*, 1994) and cellobiohydrolase I (Divne *et al.*, 1994). Sugar binding always seems to occur on the concave side of the β-sandwich, on the side of the front sheet in the legume lectins (Leonidas *et al.*, 1995; Norris *et al.*, 1994). Norris *et al.* (1994) suggested that the combination of this curved β-sheet and the connecting loops that fold over the sheet is especially suited for carbohydrate binding. The similarity between the topology of the legume lectins, the pentraxins, the galectins and the spermadhesins which have no significant sequential identity may thus be the result of convergent evolution. Moreover, the β-sandwich architecture may also be especially suited for generating a variety of quaternary structures, which is important for binding of multivalent epitopes and cross-linking of ligands (see below).

1.4 Sugar binding by legume lectins

The legume lectins have developed an ingenious framework for binding specific mono- and oligosaccharides. This framework consists of a conserved monosaccharide binding site in which four conserved residues con-

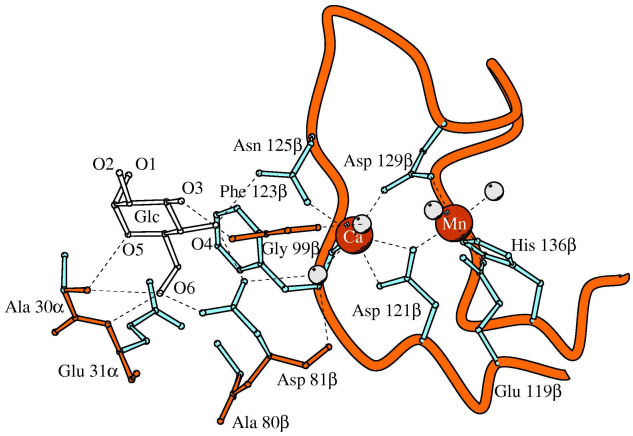


Figure 1.3: The monosaccharide and metal binding sites of lentil lectin (PDB entry 1LES), with a glucose bound in the monosaccharide binding site. The two metal ions (a Ca^{2+} ion and a transition metal ion, here shown as a Mn^{2+} ion) are ligated by six residues. The Ca^{2+} ion interacts with the four sugar binding residues : Phe 123β and Asn 125β are direct ligands, while Asp 81β and Gly 99β interact via a water molecule (shown as light gray spheres). Only the monosaccharide specificity loop (Ala 30α-Glu 31α) differs substantially between lectins belonging to different specificity groups.

fer affinity, a variable loop that confers monosaccharide specificity and a number of subsites around the monosaccharide binding site that harbor additional sugar residues or hydrophobic groups.

The legume lectins can be subdivided in five groups according to the specificity of the conserved monosaccharide binding site (Sharon & Lis, 1990) : Fuc specific, GlcNAc specific, Glc/Man specific, Gal/GalNAc specific, and the complex specificity group. Members of the latter group do not bind any simple monosaccharides. Inside these groups, some idiosyncratic representatives may be found, e.g. DBL which shows a very high preference for GalNAc over Gal (Etzler & Kabat, 1970). At present, structural data are only available for members of the latter three groups. In the monosaccharide binding site (Fig. 1.3), three of the four conserved residues (the Gly-Asn-Asp triad) hydrogen bond to the sugar, while a fourth aromatic residue stacks against the hydrophobic part of the sugar.

In some legume lectins this aromatic residue is replaced by a Leu residue, and this work shows that this has an important impact on the specificity of the lectin (Chapter 4). The four conserved residues are held in the correct position by two bound metal ions (a Ca^{2+} ion and a transition metal ion), and ligate the Ca^{2+} ion either directly or via a water molecule. Most importantly, the metal ions stabilise an unusual *cis*-peptide bond between an Ala and an Asp residue, thereby positioning the Asp side chain in the correct position for sugar binding (Bouckaert *et al.*, 1995). Gal on the one hand, and Glc and Man on the other hand can interact favourably with the four conserved residues, but must adopt different orientations relative to these residues due to their different conformations at O4 (axial for Gal, equatorial for Glc and Man). Hence, in Glc/Man-specific legume lectins like the *Vicieae* lectins or Con A, O4 and O6 hydrogen bond to the Asp residue (Bourne et al., 1990; Loris *et al.*, 1994; Naismith *et al.*, 1994),

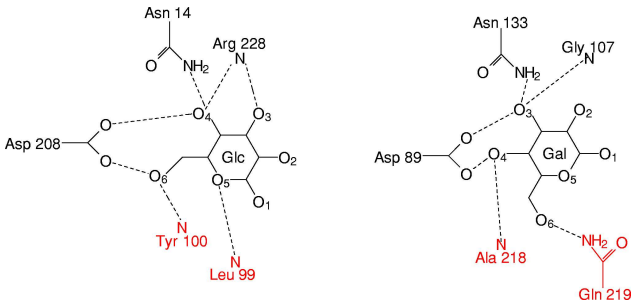


Figure 1.4: Monosaccharide recognition by Gal specific legume lectins (right, EcorL) and Man/Glc specific legume lectins (left, Con A). The residues that belong to the specificity loop, i.e., that belong to the variable part of the monosaccharide binding site, are shown in red.

while in Gal specific legume lectins like EcorL, SBA, PNA or WBAI O3 and O4 hydrogen bond (Shaanan *et al.*, 1991; Dessen *et al.*, 1995; Banerjee *et al.*, 1996; Prabu *et al.*, 1998) (Fig. 1.4). Which monosaccharide will actually bind (Gal or Glc/Man) is determined by the so-called specificity loop (Ala 30α-Glu 31α in the Glc/Man specific lentil lectin, see Fig. 1.3 and 1.4), by making it sterically impossible for the other sugars to position themselves favourably relative to the four conserved residues, and by providing additional affinity for the correct sugar (Fig. 1.5) (Sharma & Surolia, 1997). A similar combination of conserved and variable regions is used by members of the C-type lectin family to distinguish between Gal and Man/Fuc/Glc (Weis *et al.*, 1992; Kolatkar & Weis, 1996). It is expected that this scheme will also be valid for the Fuc and GlcNAc specific legume lectins, for which structural data are still lacking. Homology modeling of the fucose specific *Ulex europaeus* agglutinin I and docking of Fuc in the modeled monosaccharide binding site resulted in four possible binding modes, either involving O4 and O3, or O2 and O3 (Gohier *et al.*, 1996). Unequivocal elucidation of its mode of binding thus has to await the structure determination of a sugar complex.

Next to the monosaccharide binding site, subsites may be present that harbor additional ligands like hydrophobic aglycons or specific monosaccharides. The interactions these subsites provide are too weak to bind these groups independently, and only when attached to a central sugar residue bound in the monosaccharide binding site their role becomes important. At present, three different subsites have been described. The exact specificity of these subsites may vary from lectin to lectin. The first subsite binds hydrophobic groups (Con A; Kanellopoulos *et al.*, 1996), the second binds both monosaccharides and hydrophobic groups (Fuc in EcorL, Moreno *et al.*, 1997; and GS4, Delbaere *et al.*, 1989; Delbaere *et al.*, 1990; Delbaere *et al.*, 1993), hydrophobic groups in the *Vicieae* lectins (Kornfeld *et al.*, 1981; Loris *et al.*, 1994; Bourne *et al.*, 1994), the N-acetylgroup of GalNAc in DBL (Imberty *et al.*, 1994; Chapter 4) and the third binds monosaccharides (Fuc in the *Vicieae* lectins; Bourne *et al.*, 1994; Sokolowski *et al.*, 1997; Man in Con A; Naismith & Field, 1996; Loris *et al.*, 1996).

Only two structures are available from members of the complex specificity group, namely GS4 (Delbaere *et al.*, 1989; Delbaere *et al.*, 1990; Delbaere *et al.*, 1993) and PHA-L (Chapter 2). In the case of the GS4

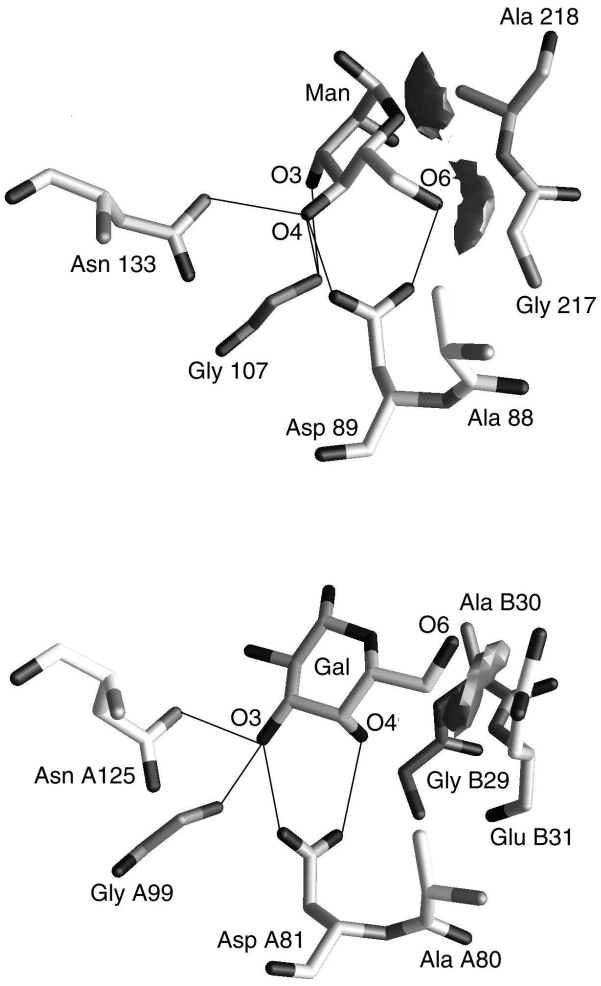


Figure 1.5: The role of the specificity loop in determining monosaccharide specificity. (Left) A Gal residue positioned in the monosaccharide binding site of a Man specific legume lectin (*Lathyrus ochrus* lectin, LOL). This was done by calculating the best least-squares superposition of the four conserved monosaccharide binding residues of the Gal specific EcorL (Asp 89, Gly 107, Phe 131 and Asn 133) on the corresponding residues of LOL. The resulting transformation was then applied to the Gal residue in the EcorL structure. The resulting clashes between the monosaccharide and the lectin were visualized with the program GRASP (Nicholls *et al.*, 1991). (Right) Similar view of Man from LOL docked in the monosaccharide binding site of the Gal specific EcorL lectin.

complexes with the Le^Y (Fuc(α1–2)Gal(β1–4)[Fuc(α1–3)]GlcNAc) and Le^b (Fuc(α1–2)Gal(β1–3)[Fuc(α1–4)]GlcNAc) oligosaccharides are available. Like in the Gal-specific legume lectins, a Gal residue is bound in the monosaccharide binding site of GS4 and interacts similarly with the four conserved residues. However, the specificity loop of GS4 adopts a distinctly different conformation and instead interacts with the (α1–4) or (α1–3) linked Fuc residue. In other words, the interactions between the specificity loop and the monosaccharide in the monosaccharide binding site have been replaced by interactions between the specificity loop and a secondary sugar residue, which explains why GS4 does not bind monosaccharides. PHA-L on the other hand does not have a "truncated" specificity loop. PHA-L probably fails to bind monosaccharides because the conserved aromatic residue in the monosaccharide binding site is replaced by a Leu residue (Chapter 2).

1.5 Quaternary structure

One of the unique features of the legume lectins is their variable quaternary structure : although the structures of their monomers are highly similar, they can associate in a number of different tetramers and dimers. In 1994 when this work was started two different tetramer types and three different dimer types were known (Fig. 1.6). This work adds a third tetramer type (PHA-L; Chapter 2 and DBL; Chapter 4) and a fourth dimer type (DB58; Chapter 4).

The so-called canonical dimer consists of two monomers that associate via the formation of a continuous twelve strand β-sheet out of two flat back sheets, along the length of the dimer. This quaternary structure is adopted by pea lectin (Einspahr *et al.*, 1986), favin (Reeke & Becker, 1986), lentil lectin (Loris *et al.*, 1993), the *Lathyrus ochrys* isolectins (Bourne *et al.*, 1994) and the α-amylase inhibitor from *Phaseolus vulgaris* (Bompard-Gilles *et al.*, 1996), which is a truncated lectin (see below). In the GS4 dimer (Delbaere *et al.*, 1989), two flat back sheets pack face to face with their b-strands running perpendicular to each other. In the EcorL dimer (Shaanan *et al.*, 1991), the dimer interface consists mainly of side chains belonging to the two upper b-strands of the back sheet. The same quaternary structure is adopted by WBA-I from winged bean (Prabu *et al.*, 1998).

Con A, the first legume lectin structure reported (Becker *et al.*, 1975), consists of two canonical dimers that pack against each other with the central parts of their continuous twelve strand β-sheets. The b-strands of the two dimers are at an angle of almost 90°, and the interface consists mainly of electrostatic interactions. The Con A tetramer has been crystallized in different space groups. These structures show that the Con A tetramer is quite flexible : small rotations and/or translation of the subunits are possible (Naismith et al., 1994; Bouckaert *et al.*, 1996). This is also illustrated by a naturally occuring "mutant" of Con A from *Canavalia brasiliensis*, whose quaternary structure differs slightly from the other Con A structures, presumably due to a mutation (Asp58Gly) at the dimer-dimer interface (Sanz-Aparicio *et al.*, 1997). Peanut agglutinin (Banerjee *et al.*, 1994; Banerjee *et al.*, 1996) forms a tetramer that can be considered as an association of two GS4 dimers. These two GS4 dimers associate via the formation of a canonical dimer, formed by one monomer from each GS4 dimer. The peanut tetramer is unique because it is the only known homotetramer that does not possess 222 nor 4 fold symmetry.

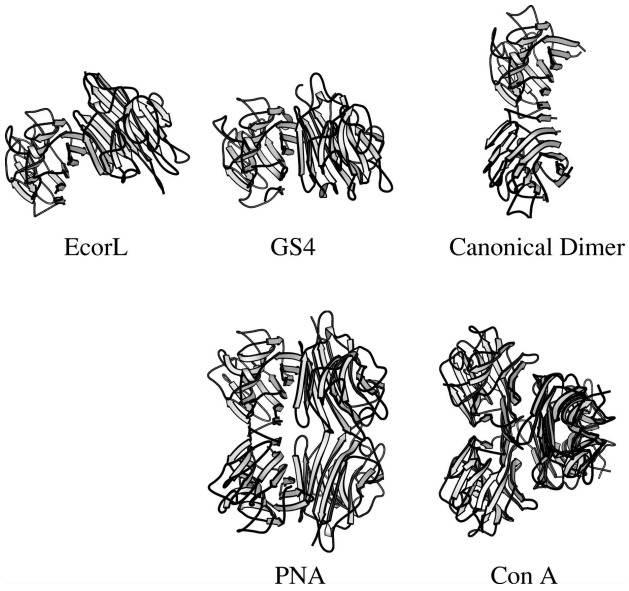


Figure 1.6: The different quaternary structures of the legume lectins. When this work was started (1994), two different tetramer types (Con A and PNA) and three different dimer types were known (the canonical dimer, GS4 and Ecorl). This work adds a fourth dimer type (DB58, Chapter 4) and a third tetramer type (PHA-L, Chapter 2, and DBL, chapter 4). The known legume lectin tetramers can be considered as dimers of dimers, so the tetramers and their corresponding dimers are vertically arranged for ease of comparison. Con A contains two canonical dimers. Peanut agglutinin contains one canonical dimer and two GS4 dimers. There is no tetramer known that ocontains the EcorL lectin type dimer.

Finally, the DBL tetramer (Chapter 2) can be viewed in two different ways : as a dimer of two canonical dimers, or as a dimer of two DB58 dimers. This mode of association creates a big channel in the center of the molecule, twice as large as the channel in DB58. In this channel, two C-terminal stretches are positioned (Dessen *et al.*, 1995), while the two other C-terminal stretches are truncated (Young *et al.*, 1995). In addition, this channel also harbors the adenine binding site found in a number of tetrameric and one dimeric legume lectin. The DBL tetramer type can probably be called the canonical tetramer type, because it is adopted by PHA-L (Chapter 2), SBA (Dessen *et al.*, 1995), DBL (Chapter 4), *Vicia villosa* isolectin B4 (Osinaga *et al.*, 1997) and *Ulex europaeus* lectin II (R. Loris, unpublished results).

It is remarkable that three of the four dimer types are also found in the tetramers : the canonical dimer is found in the Con A, DBL and peanut agglutinin tetramers; the GS4 dimer is found in peanut agglutinin, and the DB58 dimer is found in DBL. Only the EcorL dimer is not (yet?) found in any tetramer type.

Remarkably, lectins that have no significant sequence identity with the legume lectins can form similar multimers (Fig. 1.7). The galectins are Gal-specific mammalian lectins that possess the same fold as the legume lectins, although the position of the sugar binding site differs from the legume lectins. The galectin dimers of a human galectin (Lobsanov *et al.*,

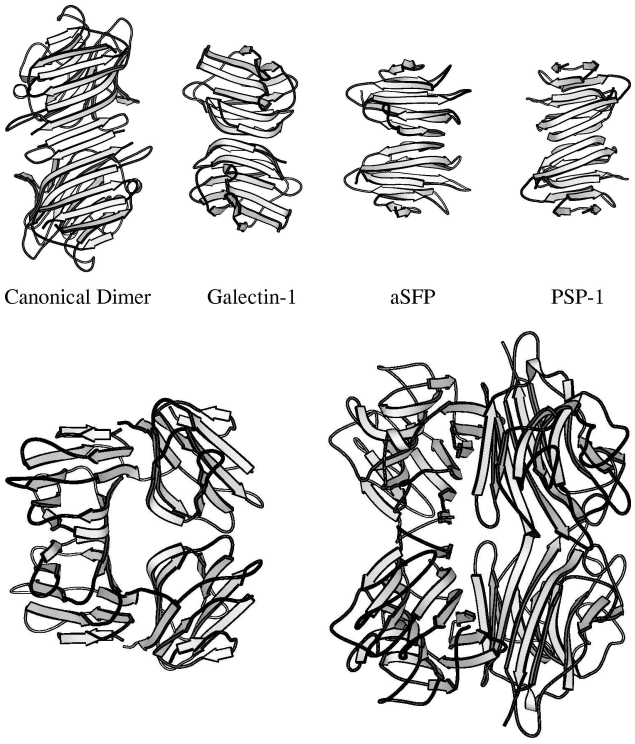


Figure 1.7: (top) From left to right.: the legume lectin canonical dimer (lentil lectin); the galectin dimer (galectin-1); the crystallographic spermadhesin dimer formed by two acidic seminal fluid protein monomers (aSFP); the dimer formed in the crystallographic porcine seminal plasma proteins I and II tetramer (PSP-I/PSP-II). Note that in the aSFP dimer the twofold axis lies parallel to the β-sheets, while in the other dimers the twofold axis is perpendicular to the β-sheets. (bottom) The PNA tetramer (right) and the PSP-I/PSP-II tetramer (left). The two canonical dimers present in both tetramers are shown in the same orientation.

1993) and bovine spleen galectin (Liao *et al.*, 1994) resemble a canonical legume lectin dimer, although two continuous β-sheets are running along the length of the dimer instead of one.

Recently, the crystal structures of two lectins from the spermadhesin family, which is involved in sperm-egg binding in mammalia, were published (aSFP and PSP-I/PSP-II; Romero *et al.*, 1997; Romao *et al.*, 1997; Varela *et al.*, 1997). Their subunits are highly similar and consist of a β-sandwich with a jellyroll related topology that contains two five stranded β-sheets, an architecture that is also found in the legume lectins and the galectins but with a different topology. Remarkably, the monomeric aSFP forms a canonical dimer in the crystal (Romao *et al.*, 1997), while the PSP-I/PSP-II heterodimer forms a tetramer in the crystal via the formation of what resembles a canonical dimer (Varela *et al.*, 1997). This tetramer bears a striking resemblance to the PNA tetramer. The PSP-I/PSP-II tetramer is observed in a trigonal crystal form of PSP-I/PSP-II, but not in a hexagonal crystal form. It is not known whether the (crystallographic) aSFP dimer and the (crystallographic) PSP-I/PSP-II tetramer are present in solution or fulfill some physiological role.

A similar phenomenon is observed for human galectin-7 (Leonidas *et al.*, 1998). In the crystal, this monomeric galectin associates into dimers that again resemble EcorL dimers. Moreover, the same dimer is observed in two different crystal forms belonging to two different space groups (C2 and $P2_12_12_1$). Again the biological relevance of this dimer remains murky.

What is the reason behind this remarkable variability of quaternary structures in the legume lectins? Multivalent lectins have been shown to form cross-linked lattices in vitro, sometimes even macroscopic crystals (Bhattacharyya *et al.*, 1989; Gupta & Brewer, 1994; Gupta et al., 1994; Brewer, 1996; Dessen *et al.*, 1995; Olsen *et al.*, 1997), with oligosaccharides that contain multiple epitopes. *In vivo*, cross linking of receptors is an important way of transferring signals (Heldin, 1995; Livnah *et al.*, 1996). Two lectins with an identical specificity, but with a different quaternary structure will necessarily form different cross-linked lattices with the same epitope. Hence, quaternary structure can be used as a tool to obtain the correct cross-liked lattice or to maximize the avidity towards structures bearing multiple epitopes.

1.6 Hydrophobic binding sites in legume lectins

Some legume lectins also possess hydrophobic binding sites that do not interfere with the sugar binding sites (Roberts & Goldstein, 1983a; Roberts & Goldstein, 1982). The function of these binding sites *in vivo* remains obscure, but the fact that these sites are present in legume lectins from different species suggests that they play a specific biological role. One of these binding sites shows a high affinity (K_a=10^6-10^5 M^{-1}) for adenine and certain adenine derivatives, including the cytokinin plant hormones (Roberts *et al.*, 1986). This site has been found in tetrameric legume lectins (PHA-L, PHA-E, LBL, SBA, (Roberts & Goldstein, 1983b; Roberts *et al.*, 1986; Roberts & Goldstein, 1982), hog peanut lectin (*Amphicarpaea bracteata*) (Maliarik *et al.*, 1987; and DBL (Roberts & Goldstein, 1983b; Gegg *et al.*, 1992; Gegg & Etzler, 1994) and a dimeric legume lectin (DB58; Gegg *et al.*, 1992; Gegg & Etzler, 1994). Binding studies suggest that one or two sites per molecule are present in tetrameric adenine binding lectins. The residues involved in the formation of the adenine binding site of PHA-E and LBL have been determined by photoaffinity labelling with 8-azidoadenine (Maliarik & Goldstein, 1988). The two isolated photolabeled peptides had an overlapping stretch of five residues in common (V^{165}LITY in PHA-E). This stretch is found in a β-sheet of the flat back sheet and lines the central channel in the PHA-L and SBA tetramers, in the vicinity of the dimer-dimer interfaces (Chapter 2). Because of the 222 symmetry of these tetramers, two stretches face each other, thereby creating a pocket with a 2-fold symmetry. Chapter 4 describes the adenine binding site of DBL in detail and confirms this hypothesis.

1.7 Truncated legume lectins

The arcelins and α-amylase inhibitors (aAI) (Mirkov *et al.*, 1994) from the seeds of the common bean are both legume lectin homologues, but they do not bind sugars and hence are not lectins. Both proteins protect the bean seeds against predation by insects, and are potentially interesting for crop protection via genetic engineering (Chrispeels, 1997). Arcelin is present

in certain wild bean strains and confers resistance against the bean pest *Zabrotes subfasciatus* (Osborn *et al.*, 1988), although it is not known how. The structure of a member of the arcelin family, arcelin-5, is described in Chapter 3. αAI inhibits α-amylase from mammals and insects, but not from bacteria, fungi or plants, and transgenic pea plants expressing αAI become resistant against certain Bruchid beetles (Shade *et al.*, 1994).

A complex of the αAI with porcine pancreas α-amylase was solved at a resolution of 1.8 Å (Bompard-Gilles *et al.*, 1996). The αAI monomer is similar to a true legume lectin monomer, although two sugar binding loops are truncated. The first truncated loop corresponds to the truncated metal binding loop in arcelin-5, while the second loop is the large W-loop that contains the Gly residue of the conserved Gly-Asn-Asp triad. Two αAI monomers form a canonical dimer, and each subunit mainly interacts with the active site of one α-amylase molecule via two loops. The structure of these two loops is stabilized by a Ca^{2+} ion, and differs from the equivalent regions in the legume lectins. The second loop corresponds to the monosaccharide specificity loop in the legume lectins.

Truncated lectins are also found outside the legume lectin family. Lithostatine (Bertrand *et al.*, 1996), a human pancreatic inhibitor of stone formation, blood coagulation factors IX/X binding protein (Mizuno *et al.*, 1997), a snake venom component, and fish type II antifreeze proteins (Gronwald *et al.*, 1998) are truncated C-type lectins. The ice binding site in the type II antifreeze proteins corresponds to the sugar binding site of a true C-type lectin (Ewart *et al.*, 1998). Charcot-Leyden crystal protein (Leonidas *et al.*, 1995), found in human eosinophils, is a truncated galectin with lysophospholipase activity.

1.8 Legume lectin structures: an overview

At present, the crystal structures of twelve native or sugar complexed legume lectins have been determined. Five belong to the Man/Glc specificity group : favin from *Vicia faba* (broad bean) (Reeke & Becker, 1986), concanavalin A from *Canavalia ensiformis* (Jack bean) (Becker et al., 1975; Naismith *et al.*, 1993) and closely related lectins from *Canavalia brasiliensis* (Sanz-Aparicio *et al.*, 1997) and *Dioclea grandiflora* (Rozwarski *et al.*, 1998), lentil lectin from *Lens culinaris* (Loris *et al.*, 1993; Casset *et al.*, 1995), pea lectin from *Pisum sativum* (Rini *et al.*, 1993) and two isolectins from *Lathyrus ochrys* (yellow-flowered pea) (Bourne *et al.*, 1994; Bourne *et al.*, 1992). Five belong to the Gal/GalNAc specificity group : *Erythrina corallodendron* (coral tree) lectin (Shaanan *et al.*, 1991; Elgavish & Shaanan, 1998), soybean agglutinin (*Glycine max*) (Dessen *et al.*, 1995; Olsen *et al.*, 1997), peanut agglutinin (*Arachis hypogea*) (Banerjee *et al.*, 1994; Banerjee *et al.*, 1996), and, more recently, *Vicia villosa* (hairy winter vetch) isolectin B4 (Osinaga *et al.*, 1997) and winged bean agglutinin (*Psophocarpus tetragonolobus*) (Prabu *et al.*, 1998). Only two crystal structures of lectins from the complex specificity group have been published : phytohemagglutinin-L (PHA-L) from *Phaseolus vulgaris* (kidney bean) (Chapter 2) and *Griffonia simplicifolia* lectin IV (GS4) (Delbaere *et al.*, 1990; Delbaere *et al.*, 1989; Delbaere *et al.*, 1993). In addition, the crystal structures of arcelin-5 (Chapter 3) and an a-amylase inhibitor from the kidney bean (Bompard-Gilles *et al.*, 1996), two plant defense proteins that are truncated and hence non-sugar binding legume lectin homologues, have been determined. Structural information about the Fuc and GlcNAc spe-

cific legume lectins is still lacking, with the exception of a recent molecular modeling study of Fuc binding by *Ulex europaeus* lectin I (Gohier *et al.*, 1996).

Despite the abundance of crystal structures, structural information on specific oligosaccharide binding by legume lectins is relatively limited. Structures of only four legume lectins in complex with a specifically recognized oligosaccharide (i.e., an oligosaccharide that is bound with a considerable higher affinity than monosaccharides) have been published. These include Con A in complex with a trisaccharide (Naismith & Field, 1996) and a pentasaccharide (Moothoo & Naismith, 1998), *Lathyrus ochrys* lectin with a fucosylated (Bourne *et al.*, 1992) and a non-fucosylated complex type sugar (Bourne *et al.*, 1994), *Griffonia simplicifolia* lectin IV with the Le^b and Le^Y tetrasaccharides (Delbaere *et al.*, 1990; Delbaere *et al.*, 1989; Delbaere *et al.*, 1993), peanut agglutinin with the T-antigen disaccharide (Ravishankar *et al.*, 1997) and *Dolichos biflorus* seed lectin in complex with the Forssman disaccharide (Chapter 4).

1.9 Homologues outside the *Leguminosae*

A few years ago, it was shown that animals contain legume lectin homologues. At present four such homologues have been described : VIP36 from mouse (Fiedler & Simons, 1995), ERGIC-53 from human (Itin *et al.*, 1996), p58 from rat and p58 from *Xenopus laevis* (Lahtinen *et al.*, 1996). VIP36 and ERGIC-53 are integral membrane proteins that contain domains which have 19 and 24 % identity with the legume lectins, respectively (Fiedler & Simons, 1994). Three conserved sugar binding residues (the Asn-Gly-Asp triad) are present in both VIP36 and ERGIC-53. VIP36 is present in the Golgi apparatus, in transport vesicles and on the cell surface and is thought to be involved in glycoprotein sorting. Similarly, ERGIC-53 is present in the intermediate region between the endoplasmatic reticulum and the Golgi apparatus, and is thought to be involved in transport of glycoproteins from the endoplasmatic reticulum to the Golgi apparatus. It has been shown that VIP36 contains two Ca^{2+} binding sites, although unlike the legume lectins VIP36 does not bind Mn^{2+} (Fiedler & Simons, 1995). In addition, GalNAc could inhibit binding of VIP36 to putative endogeneous ligands in a immunofluorescence assay. ERGIC-53 is a calcium dependent mannose specific lectin, and mutational analyses have shown that at least two of the three putative sugar binding residues are important for carbohydrate binding (Itin *et al.*, 1996). Rat p58 (Lahtinen *et al.*, 1996) is a dimeric or hexadimeric membrane protein, homologous to the legume lectins, that is thought to cycle between the ER and the golgi apparatus. A similar protein was found in *Xenopus laevis*. A legume lectin homologue (Emp47p; Schröder *et al.*, 1995) with a similar location and putative function has also been found in yeast.

The legume lectins are not the only plant lectin family for which animal homologues have been found. Comitin is a 48 kD actin-binding dimer that is primarily found on Golgi and vesicle membranes and is homologous to the mannose-specific monocotyledon lectins (Jung *et al.*, 1996). Comitin is thought to attach vesicles to the cell skeleton via an actin-binding domain that binds to the skeleton on the one hand and a Man specific lectin domain that binds to exposed Man residues on the surface of the vesicles, on the other hand. Gal binding ricin B-chain homologues were found in earthworm, tubifex, leech and lugworm (Hirabayashi *et al.*, 1998).

Recently a receptor-like serine/threonine kinase from *Arabidopsis thaliana* (*Brassicaceae*) with a domain homologous to the legume lectins has also been reported (Hervé *et al.*, 1996). This receptor is a type I membrane protein with an extra-cellular legume lectin domain. However, this domain lacks the four conserved monosaccharide binding residues, so it is unclear whether it is an active lectin. The presence of legume lectin homologues in animals and in plant species outside the Leguminoseae family (*Brassicaceae*, yeast) indicates that their *in vivo* importance might be far greater than anticipated.

These exciting examples could lead to an interesting new direction in the field of "plant" lectin studies. Moreover, they point to the necessity of using a new name (e.g. L-type lectins (Drickamer & Taylor, 1998) for the so-called "legume lectin family" since homologues of this family are found in yeast, the *Brassicaceae* plant family, amphibia and mammals.

1.10 Conclusions

The (selected) topics treated in this introduction hopefully give a good impression of the tremendous amount of work that has already been done, and of the exciting secrets that have yet to be revealed. Indeed, large pieces of the puzzle are still missing. Structural data on Fuc and GlcNAc specific lectins are still lacking, and only one sugar complex of a legume lectin with complex specificity has been published (Delbaere *et al.*, 1993). Structural characterization of a homologue from outside the *Leguminosae* family (i.e. yeast, the *Brassicaceae* plant family or mammals) and comparison with its *Leguminosae* relatives could lead to interesting new insights. The next three chapters provide new insights on sugar binding, quaternary structure, adenine/plant hormone binding and insecticidal activity of legume lectins and legume lectin homologues.

1.11 References

Banerjee, R., Mande, S.C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S.K., Suguna, K., Surolia, A. & Vijayan, M. (1994) Proc. Natl. Acad. Sci. USA, 91, 227-231.

Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. & Vijayan, M. (1996) J. Mol. Biol., 259, 281-296.

Becker, J.W., Reeke, G.N., Wang, J.L., Cunningham, B.A. & Edelman, G.M. (1975) J. Biol. Chem., 250, 1513-1524.

Bertrand, J.A., Pignol, D., Bernard, J.P., Verdier, J.M., Dagorn, J.C. & Fontecilla-Camps, J.C. (1996) EMBO J. 15, 2678-2684.

Bhattacharyya, L., Khan, M.I., Fant, J. & Brewer, C.F. (1989) J. Biol. Chem., 264, 11543-11545.

Bompard-Gilles, C., Rousseau, P., Rougé, P. & Payan, F. (1996) Structure, 4, 1441-1452.

Bouckaert, J., Loris, R., Poortmans, F. & Wyns, L. (1995) Prot. Struct.

Func. Gen., 23, 510-524.

Bouckaert, J., Loris, R. & Wyns, L. (1996) Lectins Biology-Biochemistry-Clinical Biochemistry, 11, 50-60

Bourne, Y., Mazurier, J., Legrand, D., Rougé, P., Montreuil, J., Spik, G. & Cambillau, C. (1994) Structure, 2, 209-219.

Bourne, Y., Ayouba, A., Rougé, P. & Cambillau, C. (1994) J. Biol. Chem., 269, 9429-9435.

Bourne, Y., Rougé, P. & Cambillau, C. (1992) J.Biol.Chem., 267, 197-203.

Bourne, Y., Roussel, A., Frey, M., Rougé, P., Fontecilla-Camps, J.C. & Cambillau, C. (1990) Prot. Struct. Func. Gen., 8, 365-376.

Boyd, W.C. & Shapleigh, E. (1954) Science, 119, 419

Brewer, C.F. (1996) Chemtracts Biochem. Mol. Biol., 6, 165-179.

Brewin, N.J. & Kardailsky, I.V. (1997) Trends in Plant Sci., 2, 92-98.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. (1998) Acta Cryst., D54, 905-921.

Casset, F., Hamelryck, T., Loris, R., Brisson, J., Tellier, C., Dao-Thi, M., Wyns, L., Poortmans, F., Pérez, S. & Imberty, A. (1995) J. Biol. Chem., 270, 25619-25628.

Chrispeels, M.J. (1997). In Advances in insect control. (Carozzi, N. & Koziel, M. eds), pp. 139-156, Taylor & Francis. London

Chrispeels, M.J. & Raikhel, N.V. (1991) Plant Cell, 3, 1-9.

Delbaere, L.T.J., Vandonselaar, M., Prasad, L., Quail, J.W., Nikrad, P.V., Pearlstone, J.R., Carpenter, M.R., Smillie, L.B., Spohr, U. & Lemieux, R.U. (1989) Transactions ACA, 25, 65-76.

Delbaere, L.T.J., Vandonselaar, M., Prasad, L., Pearlstone, J.R., Carpenter, M.R., Smillie, L.B., Nikrad, P.V., Spohr, U. & Lemieux, R.U. (1990) Can. J. Chem., 68, 1116-1121.

Delbaere, L.T.J., Vandonselaar, M., Prasad, L., Quail, J.W., Wilson, K.S. & Dauter, Z. (1993) J. Mol. Biol., 230, 950-965.

Dessen, A., Gupta, D., Sabesan, S., Brewer, F. & Sacchettini, J.C. (1995) Biochemistry, 34, 4933-4942.

Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Petterson, G., Knowles, J.K.C., Teeri, T.T. & Jones, T.A. (1994) Science, 265, 524-528.

Drickamer, K. & Taylor, M.E. (1998) TIBS, 23, 321-324.

Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E. & Suddath, F.L. (1986) J. Biol. Chem., 261, 16518-16527.

Elgavish, S. & Shaanan, B. (1998) J. Mol. Biol., 277, 917-932.

Emsley, J., White, H.E., O'Hara, B.P., Olivia, G., Srinivasan, N., Tickle, I.J., Blundell, T.L., Pepys, M.B. & Wood, S.P. (1994) Nature, 367, 338-345.

Etzler, M.E. (1992) In Glycoconjugates. Composition, structure and function. (Allen, H.J. & Kisailus, E.C. eds), pp. 521-539, Marcel Dekker, Inc.

Etzler, M.E. & Kabat, E.A. (1970) Biochemistry, 9, 869-877.

Ewart, K.V., Li, Z., Yang, D.S.C., Fletcher, G.L. & Hew, C.L. (1998) Biochemistry, 37, 4080-4085.

Fiedler, K. & Simons, K. (1994) Cell, 77, 625-626.

Fiedler, K. & Simons, K. (1995) J. Cell Sci., 109, 271-276.

Gabius, H.J. & Gabius, S. , eds (1997) Glycosciences, status and perspectives. Chapman & Hal, Weinheim.

Gegg, C.V. & Etzler, M.E. (1994) J. Biol. Chem., 269, 5687-5692.

Gegg, C.V., Roberts, D.D., Segel, I.H. & Etzler, M.E. (1992) Biochemistry, 31, 6938-6942.

Gohier, A., Espinosa, J.F., Jimenez-Barbero, J., Carrupt, P.A., Pérez, S. & Imberty, A. (1996) J. Mol. Graph. 14, 322-327.

Gronwald, W., Loewen, M.C., Lix, B., Daugulis, A.J., Sönnichsen, F.D., Davies, P.L. & Sykes, B.D. (1998) Biochemistry, 37, 4712-4721.

Gupta, D., Bhattacharyya, L., Fant, J., Macaluso, F., Sabesan, S. & Brewer, C.F. (1994) Biochemistry, 33, 7495-7504.

Gupta, D. & Brewer, C.F. (1994) Biochemistry, 33, 5526-5530.

Heldin, C.H. (1995) Cell, 80, 213-223.

Hervé, C., Dabos, P., Galaud, J.P., Rougé, P. & Lescure, B. (1996) J. Mol. Biol., 258, 778-788.

Hirabayashi, J., Dutta, S.K. & Kasai, K. (1998) J. Biol. Chem., 273, 14450-14460.

Imberty, A., Casset, F., Gegg, C.V., Etzler, M.E. & Pérez, S. (1994) Glycoconjugate J., 11, 400-413.

Itin, C., Roche, A.C., Monsigny, M. & Hauri, H.P. (1996) Mol. Biol. Cell, 7, 483-493.

Jung, E., Fucini, P., Stewart, M., Noegel, A.A. & Schleicher, M. (1996) EMBO J., 15, 1238-1246.

Kanellopoulos, P.N., Pavlou, K., Perrakis, A., Agianan, B., Vorgias, C.E., Mavrommatis, C., Soufi, M., Tucker, P.A. & Hamodrakas, S.J. (1996) J. Struct. Biol., 116, 345-355.

Keitel, T., Simon, O., Borris, R. & Heinemann, U. (1993) Proc. Natl. Acad. Sci., 90, 5287-5291.

Kolatkar, A.R. & Weis, W.I. (1996) J. Biol. Chem., 271, 6679-6685.

Kornfeld, K., Reitman, M.L. & Kornfeld, R. (1981) J. Biol. Chem., 256, 6633-6640.

Kuhn, P., Tarentino, A.L., Plummer, T.H. & Van Roey, P. (1994) Biochemistry, 33, 11699-11706.

Lahtinen, U., Hellman, U., Wernstedt, C., Saraste, J. & Pettersson, R.F. (1996) J. Biol. Chem., 271, 4031-4037.

Leonidas, D.D., Elbert, B.L., Zhou, Z., Leffler, H., Ackerman, S.J. & Acharya, K.R. (1995) Structure, 3, 1379-1393.

Leonidas, D.D., Vatzaki, E.H., Vorum, H., Celis, J.E., Madsen, P. & Acharya, K.R. (1998) Biochemistry, 37, 13930-13940.

Liao, D., Kapadia, G., Ahmed, H., Vasta, G.R. & Herzberg, O. (1994) Proc. Natl. Acad. Sci. USA, 91, 1428-1432.

Lis, H. & Sharon, N. (1998) Chem. Rev., 98, 637-674.

Livnah, O., Stura, E.A., Johnson, D.L., Middleton, S.A., Mulcahy, L.S., Wrighton, N.C., Dower, W.J., Jolliffe, L.K. & Wilson, I.A. (1996) Science, 273, 464-471.

Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H. & Rini, J.M. (1993) J. Biol. Chem., 268, 27034-27038.

Loris, R., Casset, F., Bouckaert, J., Pletinckx, J., Dao-Thi, M., Poortmans, F., Imberty, A., Pérez, S. & Wyns, L. (1994) Glycoconjugate J., 11, 507-517.

Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998) Biochim. Biophys. Acta, 1383, 9-36.

Loris, R., Maes, D., Poortmans, F., Wyns, L. & Bouckaert, J. (1996) J. Biol. Chem., 271, 30614-30618.

Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R. & Wyns, L. (1993) Biochemistry, 32, 8772-8980.

Maliarik, M.J. & Goldstein, I.J. (1988) J. Biol. Chem., 263, 11274-11279.

Maliarik, M.J., Roberts, D.D. & Goldstein, I.J. (1987) Arch. Biochem. Bio-

phys., 255, 194-200.

Mirkov, T.E., Wahlstrom, J.M., Hagiwara, K., Finardi-Filho, F., Kjemstrup, S. & Chrispeels, M.J. (1994) Plant Mol. Biol., 26, 1103-1113.

Mizuno, H., Fujimoto, Z., Koizumi, M., Kano, H., Atoda, H. & Morita, T. (1997) Nature Struct. Biol., 4, 438-441.

Moothoo, D.N. & Naismith, J.H. (1998) Glycobiology, 8, 173-181.

Moreno, E., Teneberg, S., Adar, R., Sharon, N., Karlsson, K. & Angström, J. (1997) Biochemistry, 36, 4429-4437.

Naismith, J.H., Habash, J., Harrop, S., Helliwell, J.R., Hunter, W.N., Wan, T.C.M. & Weisgerber, S. (1993) Acta Cryst., D49, 561-571.

Naismith, J.H., Emmerich, C., Habash, J., Harrop, S.J., Helliwell, J.R., Hunter, W.N., Raftery, J., Kalb(Gilboa), A.J. & Yariv, J. (1994) Acta Cryst., D50, 847-858.

Naismith, J.H. & Field, R.A. (1996) J. Biol. Chem., 271, 972-976.

Nicholls, A., Sharp, K.A. & Honig, B. (1991) Proteins Struct. Func. Genet., 11, 281-296.

Norris, G.E., Stillman, T.J., Anderson, B.F. & Baker, E.N. (1994) Structure, 2, 1049-1059.

Olsen, L.R., Dessen, A., Gupta, D., Sabesan, S., Sacchettini, J. & Brewer, C.F. (1997) Biochemistry, 36, 15073-15080.

Osborn, T.C., Alexander, D.C., Sun, S.S.M., Cardona, C. & Bliss, F.A. (1988) Science, 240, 207-210.

Osinaga, E., Tello, D., Batthyany, C., Bianchet, M., Tavares, G., Durán, R., Cerveñansky, C., Camoin, L., Roseto, A. & Alzari, P.M. (1997) FEBS lett., 412, 190-196.

Prabu, M.M., Sankaranarayanan, R., Puri, K.D., Sharma, V., Surolia, A., Vijayan, M. & Suguna, K. (1998) J. Mol. Biol., 276, 787-796.

Ravishankar, R., Ravindran, M., Suguna, K., Surolia, A. & Vijayan, M. (1997) Current Science, 72, 855-861.

Reeke, G.N. & Becker, J.W. (1986) Science, 234, 1108-1111.

Rini, J.M. (1995) Annu. Rev. Biophys., 24, 551-577.

Rini, J.M., Hardman, K.D., Einspahr, H., Suddath, F.L. & Carver, J.P. (1993) J. Biol. Chem., 268, 10126-10132.

Roberts, D.D., Arjunan, P., Townsend, L.B. & Goldstein, I.J. (1986) Phytochemistry, 25, 589-593.

Roberts, D.D. & Goldstein, I.J. (1982) J.Biol.Chem., 257, 11274-11277.

Roberts, D.D. & Goldstein, I.J. (1983a) Arch. Biochem. Biophys., 224, 479-484.

Roberts, D.D. & Goldstein, I.J. (1983b) J. Biol. Chem., 258, 13820-13824.

Romao, M.J., Kölln, I., Dias, J.M., Carvalho, A.L., Romero, A., Varela, P.F., Sanz, L., Töpfer-Petersen, E. & Calvete, J.J. (1997) J. Mol. Biol., 274, 650-660.

Romero, A., Romao, M.J., Varela, P.F., Köln, I., Dias, J.M., Carvalho, A.L., Sanz, L., Töpfer-Petersen, E. & Calvete, J.J. (1997) Nature Struct. Biol., 4, 783-787.

Rozwarski, D.A., Swami, B.M., Brewer, C.F. & Sacchettini, J.C. (1998) J. Biol. Chem., 273, 32818-32825.

Sanz-Aparicio, J., Hermoso, J., Grangeiro, T., Calvete, J. & Cavada, B. (1997) FEBS lett., 405, 114-118.

Schröder, S., Schinmöller, F., Singer-Krüger, B. & Riezman, H. (1995) J. Cell. Biol., 131, 895-912.

Shaanan, B., Lis, H. & Sharon, N. (1991) Science, 254, 862-865.

Shade, R.E., Schroeder, H.E., Pueyo, J.J., Tabe, L.M., Murdock, L.L., Higgins, T.J.V. & Chrispeels, M.J. (1994) Bio/Technology, 12, 793-796.

Sharma, V. & Surolia, A. (1997) J. Mol. Biol., 267, 433-445.

Sharon, N. & Lis, H. (1990) FASEB J., 4, 3198-3208.

Sokolowski, T., Peters, T., Pérez, S. & Imberty, A. (1997) J. Mol. Graph., 15, 37-42.

Srinivasan, N., Rufino, S.D., Pepys, M.D., Wood, S.P. & Blundell, T.L. (1996) Chemtracts-Biochem. Mol. Biol., 6, 149-164.

Stirk, H.J., Woolfson, D.N., Hutchinson, E.G. & Thornton, J.M. (1992) FEBS lett., 308, 1-3.

Van Damme, E.J.M., Peumans, W.J., Pusztai, A. & Bardocz, S. (1998). Handbook of plant lectins: properties and biomedical applications. John Wiley & Sons, New York.

Varela, P.F., Romero, A., Sanz, L., Romao, M.J., Töpfer-Petersen, E. & Calvete, J.J. (1997) J. Mol. Biol., 274, 635-649.

Weis, W.I. & Drickamer, K. (1996) Annu. Rev. Biochem., 65, 441-473.

Weis, W.I., Drickamer, K. & Hendrickson, W.A. (1992) Nature, 360, 127-134.

Young, M.N., Watson, D.C., Yaguchi, M., Adar, R., Arango, R., Rodriguez-Arango, E., Sharon, N., Blay, P.K.S. & Thibault, P. (1995) J. Biol. Chem., 270, 2563-2570.

Chapter 2

PHA-L, a toxic lectin from the common bean (*Phaseolus vulgaris*)

Reference: 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

2.1 Introduction

The seeds of the common bean (*Phaseolus vulgaris*) contain a protein fraction with sugar binding and hemagglutinating properties, called phytohemagglutinin (PHA) (Chrispeels & Raikhel, 1991). This fraction consists of five different legume lectins, built out of two subunits (L and E) in all possible tetrameric combinations (L_4, L_3E, L_2E_2, LE_3 and E_4). The L_4 and E_4 tetramers are called PHA-L and PHA-E, respectively. The E-type and L-type subunits are respectively responsible for the erythroagglutinating and leucoagglutinating properties of the PHA fraction. PHA-L is a potent mitogen (Licastro *et al.*, 1993) and is used as a tool in maintaining lymphocyte cultures, cancer diagnosis and cell biology (Van Damme *et al.*, 1997).

Raw or undercooked beans are toxic for birds and mammals (including humans) due to the presence of the PHA fraction. According to the U.S. Food & Drug Administration, the symptoms of PHA poisoning start three hours after consumption and include extreme nausea, vomiting and diarrhea. Recovery is spontaneous three to four hours after onset of symptoms. A classic example of large scale PHA poisoning occurred in West Berlin in 1948, when the population received improperly cooked beans during the air lift (Gabriel, 1950). PHA poisoning is not well known in the medical community, and the cause is often erroneously attributed to bacterial food poisoning. PHA binds to the epithelium of the gut and mediates binding and overgrowth of the commensal flora, which is the probable cause of the toxic effect (for a review, see Pusztai & Bardocz, 1996). Germ free rats are essentially immune for the toxic activity of PHA.

The E- and L-type subunits are both members of a family of four different polypeptides encoded by four tightly linked genes, generally referred to as the phytohemagglutinin family of bean proteins (Mirkov *et al.*, 1994). This family, in addition to PHA-E and PHA-L, also contains arcelin (Romero

Andreas *et al.*, 1986; Hartweck *et al.*, 1991; Goossens *et al.*, 1994), which exists in at least six electrophoretic forms, and an α-amylase inhibitor (aAI) (Moreno *et al.*, 1989). Arcelin and αAI can be considered as truncated legume lectins, in which respectively one and two loops that play a crucial sugar binding role are missing, abolishing the sugar binding properties. PHA-L, PHA-E, arcelin (Osborn *et al.*, 1988) and αAI (Shade *et al.*, 1994) protect bean seeds against predation although the precise origin of the toxicity of arcelin is unknown.

PHA-L is a glycoprotein : each subunit is N-glycosylated at two different sites, with consensus sequence Asn-X-Ser/Thr. The subunit possesses a high mannose type glycan attached at Asn 12, and a complex type glycan at Asn 60 (Sturm *et al.*, 1986). The highest affinity binder to PHA-L is the pentasaccharide Gal(β1–4)GlcNAc(β1–2)[Gal(β1–4)GlcNAc(β1–6)]-Man, which is found in tetra- and tri-antennary complex type oligosaccharides of mammalian origin (Hammarström *et al.*, 1982). PHA-L and PHA-E both belong to the complex specificity group of legume lectins : precipitation inhibition studies indicate that monosaccharides are not active as inhibitors. For PHA-L, the minimal oligosaccharide active as inhibitor is the disaccharide GlcNAc(β1–2)Man.

In addition to sugars, some legume lectins also bind adenine and adenine derived plant hormones, i.e., the cytokinins. High affinity binding sites (K_a=10^5-10^6 M^{-1}) for adenine and its derivatives have been found for PHA-L, PHA-E, *Dolichos biflorus* seed lectin (DBL), *Dolichos biflorus* stem and leaf lectin (DB58), soybean agglutinin (SBA), *Phaseolus lunatus* lectin (LBL) and hog peanut agglutinin (Roberts and Goldstein, 1982; Roberts and Goldstein, 1983a; Gegg *et al.*, 1992).

In this chapter the crystal structure of PHA-L, the L^4 tetramer, is presented. The putative location of the adenine/plant hormone binding site is presented, based on photoaffinity labelling of this site in PHA-E and LBL (Maliarik *et al.*, 1988). Furthermore, the structure of the conserved monosaccharide binding site is described with respect to the complex specificity of PHA-L.

2.2 Materials and methods

2.2.1 Crystallization and data collection

Crystallization of PHA-L has been described elsewhere (Dao-Thi *et al.*, 1995). Briefly, the crystals were grown at 4°C by vapor diffusion using the hanging drop method. The composition of the bottom solution was 8% (w/v) PEG 6 kDa (Janssen Pharmaceutical) and 100 mM Tris, pH 8.5. Droplets consisted of 5 μl bottom solution and 5 μl 5mg/ml PHA-L solution. PHA-L was purchased from Sigma. PHA-L crystallized in the monoclinic space group C2, with cell parameters a=106.3 Å, b=121.2 Å, c=90.8 Å and b=93.7°. The asymmetric unit contains one complete PHA-L tetramer. Data collection was performed on a single crystal at room temperature with an Enraff-Nonius FAST area detector, using Cu Kα radiation generated by a rotating anode X-ray generator (operating at 40 kV, 98 mA). Data reduction was done using the MADNESS software package (Pflugrath and Messerschmidt, 1989). The statistics of the data collection are given in tables 2.1 and 2.2.

Table 2.1: Data collection statistics of the native PHA crystal.

Resolution (Å)	Observed reflection	Unique reflections	Completeness (%)	Multiplicity	Rmerge
6.78-10.00	2310	1327	96.0	1.7	0.051
5.46-6.78	4766	1777	97.6	2.7	0.080
4.70-5.46	7746	2094	96.7	3.7	0.079
4.19-4.70	8649	2384	96.8	3.6	0.090
3.81-4.19	7362	2581	94.7	2.9	0.099
3.52-3.81	5967	2702	91.0	2.2	0.081
3.29-3.52	6193	2874	90.2	2.2	0.111
3.10-3.29	6159	2974	87.6	2.1	0.149
2.94-3.10	6217	3052	85.1	2.0	0.209
2.80-2.94	6176	3085	81.8	2.0	0.293
Total	61545	25850	88.5	2.5	0.097

2.2.2 Structure determination and interpretation

All calculations were done on a Silicon Graphics INDY workstation and on a Cray YMP supercomputer. Molecular replacement was carried out with the AMORE software package (Navaza, 1994). The structure of the lentil lectin dimer complexed with sucrose, solved at a resolution of 1.9 Å (Casset *et al.*, 1995; entry 1LES in the Brookhaven database) was used as search model, after removal of the solvent molecules, the metal ions and the two bound sucrose molecules. Only two clear solutions were found that were consistent with a tetramer in the asymmetric unit while showing reasonable contacts with the symmetry mates. An initial rigid body refinement was performed using X-PLOR (Brünger, 1992), using data between 10.0 and 4.0 Å, which caused a drop of the R-factor from 52.5 % to 41.0 %. Model building was done with O (Jones *et al.*, 1991), by making extensive use of the provided library of protein structures. Further refinement consisted of restrained individual B value refinement, POWELL positional refinement and simulated annealing (4000 to 300 K, Δt=0.0001 ps) with X-PLOR. All data between 10 and 2.8 Å were used. During the refinement, non-crystallographic symmetry restraints between the four monomers were applied (weight=300.0 kcal mol^{-1} Å-2, σ=1.0 Å^2). After each refinement step the quality of the model was checked using PROCHECK (Laskowski *et al.*, 1993). It was decided not to systematically add water molecules to the structure, as they can cause considerable model overfitting at 2.8 Å resolution. An exception was made for the four water molecules which are ligands of the Ca^{2+} and Mn^{2+} ions present in each monomer, since the electron density clearly justified this and since they are a very conserved feature of the legume lectin crystal structures (Loris *et al.*, 1994). The final R factor of the converged structure was 20.0 % for 24850 reflections between 10 and 2.8 Å. The a posteriori free R value (Brünger, 1993), calculated after subjecting the final structure to simulated annealing refinement against the working set, was 22.9 %. Hydrogen bonds were calculated using the HBPLUS program (McDonald *et al.*, 1993). Entry 1SBA (Dessen *et al.*, 1995) in the Brookhaven database was used to compare PHA-L with SBA. Accessible surface areas were calculated with the GRASP program (Nichols *et al.*, 1991).

Table 2.2: Overview of the statistics of data collection and refinement.

Space group	C2
Unit cell	a=106.3 Å , b=121.2 Å, c=90.8 Å β=93.7 °
Resolution	2.8 Å
Number of observations	61545
Number of unique refelections	25850
Multiplicity	2.5
Completeness (2.80 Å-2.94 Å, last shell)	88.5 % (81.8 %)
Rmerge (2.80 Å-2.94 Å, last shell)	9.7 % (29.3 %)
Final R-value (10.0 Å-2.8 Å)	20.0 %
A posteriori free R value (10.0 Å-2.8 Å)	22.9 %
Number of protein atoms	3838
Number of solvent atoms	16
Number of sugar atoms	46
Number of metal ions	4 Mn, 4 Ca
Rms on bond lengths	0.017 Å
Rms on bond angles	2.01°
Rms on dihedrals	28.16°
Rms on impropers	1.71°
Distribution of non-Gly non-Pro $\phi - \psi$ angles in the Ramachandran plot	85.8 % in core regions 13.7 % in additionally allowed regions 0.5 % in generously allowed regions 0.0 % in disallowed regions

2.3 Results and discussion

2.3.1 Structure description

In accordance with the currently accepted refinement protocols, the structure has been refined using non-crystallographic symmetry restraints between the four subunits (termed A, B, C and D). Hence, the structures of the four subunits are virtually identical, with the exception of a few regions in which the quality of the electron density varies between the different subunits (see below). The rms differences between the α-carbon positions of the four subunits, calculated with the lsq_explicit command of the O program, vary between 0.11 and 0.16 Å.

The structure of the PHA-L monomer contains 233 residues of the 252 residues of the complete monomer, based on the sequence of the PHA-L encoding gene (Hoffman *et al.*, 1985) : no electron density is observed for the final 19 C-terminal residues. The probable reasons for this absence are discussed below.

Of the two covalently attached oligosaccharides per monomer (Sturm et al., 1986) only the core GlcNAc residue of the high mannose glycan attached to Asn 12 is clearly visible in the electron density, probably due to the flexibility of these glycan moieties. No interpretable density is observed for the small complex type glycan attached to Asn 60.

All residues are located in the allowed regions of the Ramachandran plot. A number of regions located in surface exposed loops suffer from

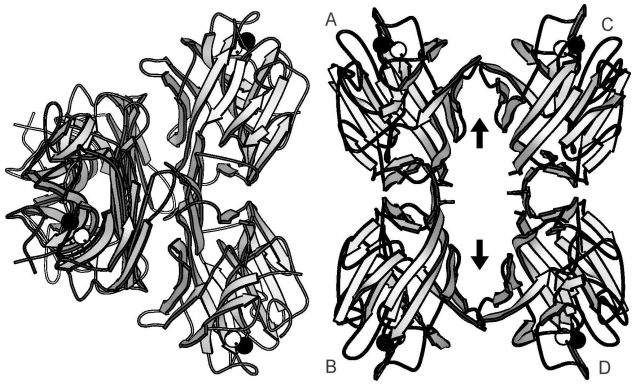


Figure 2.1: Comparison between Con A and PHA-L. The PHA-L tetramer is shown on the right, the Con A tetramer is shown on the left. The right dimers in both tetramers have the same orientation to emphasize the difference in dimer-dimer packing between PHA-L and Con A. The central channel running between the two dimers in PHA-L is clearly visible. The putative position of the adenine binding sites is indicated with arrows. All figures in this chapter were made using MOLSCRIPT (Kraulis, 1991).

rather poor electron density or breaks in the density for the main chain. Breaks in the density of the main chain were observed for regions A36-A38, C36-C38 and D35-D38. These regions were consequently removed from the model. Monomer B shows the best electron density : for this monomer, no main chain electron density breaks are observed, and the overall quality of the electron density is slightly higher. Like the other solved legume lectin structures, the PHA-L subunit consists of a flat, six-stranded β-sheet, and a curved, seven-stranded β-sheet.

2.3.2 Quaternary structure

PHA-L is a tetramer with approximate dimensions 40 Å $\times$ 60 Å $\times$ 80 Å, consisting of four identical subunits (Fig. 2.1).

Each monomer is involved in two different monomer-monomer interfaces. The first interface is a conventional β-sheet like contact between two strands, creating a continuous, curved, antiparallel twelve strand β-sheet spanning two monomers (interfaces A-B and C-D). The other interface is mainly formed by van der Waals interactions between two β-strands (interfaces A-C and B-D) (Fig. 2.2a and 2.2b). In the tetramer, the two curved 12-strand β-sheets face each other, creating a large channel between them. The same tetrameric organisation is also found in SBA, although the SBA tetramer was prevously incorrectly described as identical to the Con A tetramer (Dessen *et al.*, 1995). As for most tetramers (Miller, 1989), the point group of the PHA-L tetramer is 222.

At present, the X-ray structures of three other legume lectin tetramers (peanut agglutinin (PNA, Banerjee *et al.*, 1994), SBA (Dessen *et al.*, 1995) and Con A (Becker *et al.*, 1975)) and of three different types of dimers (*Griffonia simplicifolia* isolectin IV (GS4, Delbaere *et al.*, 1990), *Erythrina corallodendron* lectin (EcorL, Shaanan *et al.*, 1991) and the "canonical dimer", as represented by lentil lectin and other *Vicieae* lectins (Loris *et*

Table 2.3: Van der Waals contacts between the two canonical dimers in PHA-L for the A-C dimer interface. Atom distances below 4 Å are considered as belonging to atoms involved in van der Waals contacts. The B-D interface is virtually identical.

Monomer A	Monomer C	Number of contacts
Lys 149	Lys 184	1
Leu 173	Val 179	1
Val 179	Leu 173	1
Pro 181	Leu 173	1
Lys 184	Lys 149, Ser 190, Asp 191, Thr 192	11
Thr 185	Ser 190	2
Ser 186	Ile 188, Val 189, Ser 190	9
Phe 187	Ile 188	1
Ile 188	Ser 186, Phe 187, Ile 188	9
Val 189	Ser 186	1
Ser 190	Lys 184, Thr 185, Ser 186	10
Asp 191	Lys 184	1
Thr 192	Lys 184	2

al., 1993) are known. The structures of the tetramers solved so far can be described as classic "dimers of dimers" with a 222 symmetry, with the exception of PNA. The PNA tetramer is a dimer of two GS4-like dimers, but unlike Con A, SBA and PHA-L, this protein does not posses 222 symmetry. Although SBA and PHA-L on the one hand, and Con A on the other hand both posses 222 symmetry and both consist of two canonical dimers (dimers A-B and C-D in PHA-L), their quaternary structures are clearly different (Fig. 2.1). Con A consists of two canonical dimers that pack together through the formation of a dimer/dimer interface that involves the central part of both canonical dimers. The contact between the two dimers consists mainly of loop interactions. The two canonical dimers in PHA-L pack together mainly through close contacts between the two outmost strands of the 12 strand β-sheet of both dimers. The contacts mainly involve van der Waals interactions between the side chains of these two strands, from residues 181 to 192. Since the dimer interface consists of two strands in two β-sheets that face each other, the side chains of the residues in these strands are consecutively oriented towards the inside of the monomer and towards the interface. Thus, the residues that mostly contribute to the van der Waals contact surface are residues Ser 186, Ile 188 and Ser 190 while Thr 185, Phe 187, Val 189 and Asp 191 only contribute marginally (see table 2.3). The fact that the protruding side chains are small (Ile and Ser), allows them to be intercalated in a "zipper like" fashion (Fig. 2.2a and Fig 2.2b).

Intercalation of side chains is normally not present in β-sheet face to face packing, where the side chains of the two sheets involved can be separated by a plane and the interface is essentially smooth (Chothia and Janin, 1981). Intercalation of side chains in β-sheet packing was first proposed for silk β-fibroin, and has been observed in pyruvoyl-dependent histidine decarboxylase from *Lactobacillus* (Gallagher *et al.*, 1993; Murzin, 1994). The intercalation of the side chains also allows the two strands in a

Table 2.4: Hydrogen bonds between the two canonical dimers in the PHA-L tetramer.

Dimer 1	Dimer 2	Distance (Å)
Lys 184A O	Ser 190C Oγ	2.9
Ser 186A Oγ	Ser 190C Oγ	2.7
Ser 190A Oγ	Lys 184C O	3.0
Ser 190A Oγ	Ser 186C Oγ	2.6
Lys 184B O	Ser 190D Oγ	2.8
Ser 186B Oγ	Ser 190D Oγ	2.8
Ser 190B Oγ	Lys 184D O	3.1
Ser 190B Oγ	Ser 186C Oγ	2.8

silk β-sandwich to be closer to each other than in the default packing : the backbone atoms are only 6 Å apart in pyruvoyl-dependent histidine decarboxylase, as opposed to 10 Å in a classic β-sheet. The backbone atoms of the two interacting β-strands in PHA-L are 6.5 Å apart. Eight hydrogen bonds are made in the interface between the two dimers, involving two serine side chain hydroxyl groups (Ser 186, Ser 190) and a lysine main chain oxygen atom (Lys 184) in each monomer (see table 2.4).

Lys 184 lies on the end of the β-strand and positions its side chain almost in parallel with the β-strand, making van der Waals contacts with three residues from the other strand (Ser 190, Asp 191, Thr 192).

The way in which the two dimers pack in PHA-L creates a large channel in the centre of the molecule : the two twelve strand β-sheets facing each other are approximately 18 Å apart in the centre. Similar channels are found in tetramers where the monomers, as in PHA-L and Con A, only pairwise interact (Miller, 1989).

2.3.3 The metal and sugar binding sites

All legume lectins posses two bound metal ions (one Ca^{2+} ion and one transition metal ion, mainly Mn^{2+}) per monomer in the vicinity of the sugar binding site. The presence of these two bound metal ions is vital for the sugar binding capabilities of the legume lectins. The structure of the metal binding site is extremely conserved in the legume lectin structures solved up to now. The two metal ions are ligated by four water molecules and six amino acid residues : one His, one Glu, one Asn, two Asp and one hydrophobic residue (Phe in lentil lectin; Loris *et al.*, 1993; Casset *et al.*, 1995; pea lectin; Rini *et al.*, 1993; the *Lathyrus ochrys* isolectins; Bourne *et al.*, 1990; Bourne *et al.*, 1994; EcoRL, SBA; Tyr in Con A and PNA; Trp in GS4). As expected, the metal binding site of PHA-L is highly similar to the metal binding sites found in the other legume lectin structures. The Mn^{2+} ion is co-ordinated by His 137, Glu 122, Asp 124 and Asp 132. The Ca^{2+} ion is co-ordinated by Leu 126, Asp 124, Asn 128 and Asp 132. The Ca^{2+} ion interacts via a water molecule with an oxygen atom of the side chain of Asp 86, and stabilizes the conserved Ala 85 Asp 86 *cis*-peptide bond, that is present in all solved legume lectin structures.

In all the other legume lectin structures, the position of Leu 126 is occupied by a residue with an aromatic side chain (Phe, Tyr or Trp). This aromatic residue stacks against the sugar ring and is an important feature

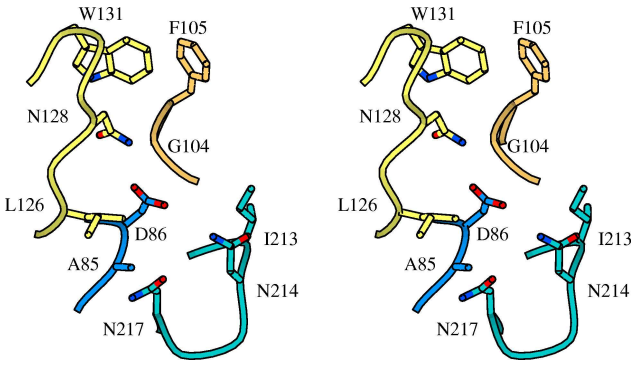


Figure 2.2: The sugar binding site of PHA-L. The metal binding loop is shown in yellow; the Ala-Asp *cis*-peptide bond in blue; the Ω-loop in ochre and the monosaccharide specificity loop in green. As explained in Chapter 4, Trp 131 and Phe 105 form a subsite that binds monosaccharides and hydrophobic groups.

of sugar binding by legume lectins. PHA-L belongs to the complex specificity group, i.e., it does not appreciably bind simple monosaccharides. The minimal oligosaccharide that is able to inhibit precipitation by PHA-L is the disaccharide GlcNAc(β1–2)Man (Hammarström *et al.*, 1982). Only one other crystal structure of a legume lectin with a complex specificity has been described: the structure of the *Griffonia symplicifolia* isolectin IV (GS4, Delbaere *et al.*, 1990). In GS4, one of the four sugar binding loops (the monosaccharide specificity loop) is truncated, thereby abolishing monosaccharide binding. The structure of GS4 in complex with the Le^b and LeY tetrasaccharides shows that a Gal residue is bound in the truncated monosaccharide binding site, while two Fuc residues are bound in neighboring subsites. These two Fuc residues thus compensate for the truncated sugar binding loop. Inspection of the corresponding loops in PHA-L indicates that none of the loops are truncated (Fig. 2.3). It is likely that PHA-L does not (or only weakly) bind monosaccharides due to the lack of aromatic stacking against the sugar ring. The seed lectin from *Dolichos biflorus*, which also has a Leu residue at this position (Chapter 4), uses a similar mechanism to distinguish between Gal and GalNAc.

2.3.4 Comparison between PHA-L and SBA

SBA and PHA-L share the same quaternary structure. To investigate the similarity of the interfaces of both proteins, we superimposed both dimer-dimer interfaces. The interface between the two canonical dimers in SBA is extremely similar to the PHA-L interface. As in PHA-L, the association of the two subunits occurs through intercalation of the side chain of two Ser (Ser 187 and Ser 191 in SBA) and one Ile residue (Ile 189 in SBA). The only substantial difference between the two interfaces is the conservative substitution of Lys 184 in PHA-L by Arg 185 in SBA. The long, hydrophobic part of the Arg 185 side chain in SBA is involved in similar van der Waals contacts, but also forms an additional salt bridge with the Asp 192 residue. The structure of the SBA dimer-dimer interface is shown in Fig. 2.2b.

It has been suggested that in SBA, a C-truncated subunit (240 amino-acids) and a non-truncated subunit (253 amino-acids) have to face each

other in the tetramer, because the C-terminal portion of the intact subunit positions itself between the two subunits while the space between the subunits is too small to allow for two C-terminal portions (Dessen *et al.*, 1995). C-terminal truncation has also been reported for PHA-E (Young *et al.*, 1995), but was never investigated for PHA-L. The intact PHA-E subunit consists of 254 amino-acids, the truncated subunit consists of 244 amino-acids. In PHA-L, no electron density is observed beyond 233, due to either C-terminal truncation or chain flexibility, or possibly a combination of both. Inspection of the density between the two dimers did not clearly reveal the presence of the C-terminal end of the chain in the space between the two dimers as was reported for SBA, although some uninterpretable electron density was observed.

Packing of the two canonical dimers buries about 1500 $Å^2$ of the protein surface in PHA-L, comparable to 1200 $Å^2$ for SBA.

2.3.5 Putative position of the adenine/cytokinin binding site

In addition to sugar binding sites, some legume lectins also possess hydrophobic binding sites that can be divided into three groups. The first group of sites is adjacent to the sugar binding site and is responsible for the higher affinity of these lectins for monosaccharides bearing certain hydrophobic substituents, as compared to the non substituted monosaccharides (for a review, see Loris *et al.*, 1998). The second site is positioned at a distance of approximately 30 Å from the sugar binding site and has a low affinity (10^2 M^{-1}) for hydrophobic ligands like the plant hormone indoleacetic acid, TNS and ANS (Yang *et al.*, 1974; Edelman & Wang, 1978; Kella *et al.*, 1984). Finally, the third site is a high affinity (10^6 M^{-1}) binding site for adenine and its N6 derivatives, including adenine derived plant hormones, i.e., cytokinins. This site also binds TNS, but not ANS. High affinity binding of adenine has been shown for SBA, DBL, DB58, PHA-E, PHA-L, hog peanut agglutinin and LBL (Roberts & Goldstein, 1983a,b; Maliarik *et al.*, 1987; Maliarik & Goldstein, 1988; Maliarik *et al.*, 1989; Gegg *et al.*, 1992). All these lectins are tetramers, except DB58, which is a dimer. There has been some controversy concerning the number of adenine binding sites per tetramer. A single site per tetramer has been reported for LBL and DBL (Roberts and Goldstein, 1982; Roberts and Goldstein, 1983b), but more recent research yielded two adenine binding sites, suggesting a symmetric binding site at the interface between two subunits (Gegg *et al.*, 1992).

The exact function of these hydrophobic binding sites is unknown, but legume lectins might play a role in plant hormone signalling or storage. Moreover, plant hormones are often stored in the plant under a glycosylated form (Sembdner *et al.*, 1994).

The sequences of the adenine binding sites in LBL and PHA-E have been determined by photoaffinity labelling (Maliarik *et al.*, 1988). The two determined sequences had an overlapping five residues in common, with sequence Val-Leu-Ile-Thr-Tyr (residues 165 to 169 in PHA-E, residues 164 to 168 in LBL) (Jordan *et al.*, 1994). This sequence is situated in a highly conserved region among the adenine binding legume lectins, and is also present in PHA-L. Furthermore, PHA-E and PHA-L (82 % identity) can exchange subunits and must therefore have the same quaternary structure. Investigating the position of this sequence in the PHA-L structure may

therefore give us an indication about the global position of the adenine binding site. In the PHA-L structure, this sequence is found in the third strand of the flat six strand β-sheet, from residue 163 to residue 167 (Fig. 2.1).

This β-sheet is located in the central channel present in PHA-L. If this region is indeed involved in adenine binding, then localisation of this region in the solved PHA-L structure suggests that there are two adenine binding sites per tetramer, located at both ends of the protein, in agreement with the work done by Gegg *et al.* (1992). The organisation of the PHA-L tetramer is such that the labelled five residue sequences of two monomers are close together, and are related by the two-fold axis along the length of the molecule, thus creating a binding site with a two-fold symmetry. Apparently, the adenine binding site is located near the interface between the two canonical dimers that form a tetramer, suggesting an interplay between quaternary structure and function. The two possible binding sites are located approximately 30 Å from each other.

Similar ligand binding in a symmetric binding site, created by the packing of the subunits in a tetramer, also occurs in the human serum protein transthyretin (formerly known as prealbumin). In transthyretin, two monomers associate into a dimer through the formation of a continuous antiparallel β-sheet (Blake *et al.*, 1974; Blake *et al.*, 1978) (Fig. 2.4). Two such dimers pack face to face to form a tetramer, forming a channel between the two β-sheets, as in PHA-L. This channel accommodates two identical binding sites for thyroxine, a steroid hormone (Blake and Oatley, 1977; Ciszak *et al.*, 1992; Wojtczak *et al.*, 1992). Although at first only one binding site for thyroxine was reported, further research showed that binding of thyroxine in one of the binding sites (10^8 M^{-1}) decreases the affinity of the other site (10^6 M^{-1}) (Ferguson *et al.*, 1975). Again as in PHA-L, these sites show internal two-fold symmetry. In addition, transthyretin binds ANS in the thyroxine binding sites. Although the adenine binding site of PHA-E does not bind ANS, it does bind the related ligand TNS. In both transthyretin and PHA-L, the central channel is lined with small, hydrophobic residues (Val, Ala, Ile, Leu) and Ser or Thr side chains. Inspection of the channel in SBA, which also binds adenine and TNS, revealed a very similar pattern of small, hydrophobic and hydroxyl containing side chains. These striking similarities between PHA-L and transthyretin provide further support for the model of adenine binding in legume lectins proposed here.

2.4 Conclusions

Although PHA-L belongs to the complex specificity group, it does not have truncated sugar binding loops. Rather, lack of aromatic stacking against the sugar ring is probably the reason that PHA-L does not appreciably bind monosaccharides: in the monosaccharide binding site of PHA-L, a conserved aromatic residue is replaced by an aliphatic residue (Leu 126).

PHA-L is a tetrameric protein, in which two canonical legume lectin dimers pack together through the intercalation of the side chains of the two outmost β-strands. This way of packing creates a large channel in the centre of the protein, lined mainly with Ser, Thr and small, apolar residues (Val, Ala, Ile, Leu). We suggest that this channel forms the adenine and cytokinin binding site, based on photoaffinity labelling (Maliarik and Goldstein, 1988) and comparison with transthyretin. The parallels between our

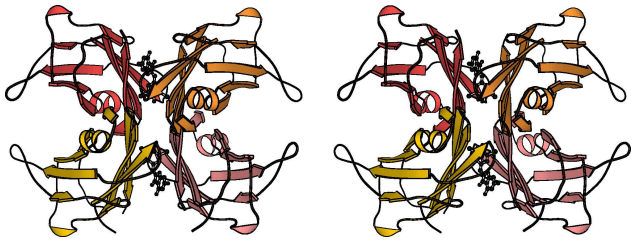


Figure 2.3: Stereo figure of the transthyretin tetramer complexed with 3-dibromo-2,4,4,6-tetrahydroxyaurone, a flavone derivative. The ligand (shown as a ball-and-stick model) is bound in the central hole that runs through the molecule, through interactions with side chains of the residues that make up the flanking β-strands (mainly Ser, Thr, Ala and Leu). The binding site possesses twofold symmetry. The central hole is approximately 10 Åwide. The four subunits are shown in different colors.

model of adenine binding by PHA-L and thyroxin binding by transthyretin are striking.

- Both consist of similar dimers.
- The packing of both dimers in the tetramer is similar.
- They both contain a large channel, lined with mainly Ser, Thr and residues with small hydrophobic side chains, created by the face to face positioning of two β-sheets.
- Both transthyretin and PHA-L bind similar ligands (ANS and TNS).

Our results indicate that PHA-L and transthyretin apparently converged towards a similar framework for binding small ligands with aromatic groups.

2.5 References

Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Surolia, A. & Vijayan, M. (1994) Proc. Natl. Acad. Sci. USA, 91, 227-231

Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1975) J. Biol. Chem. 25, 1513-1524

Blake, C. C. F., Geisow, M. J., Oatley, S. J., Rerat, B. & Rerat, C. (1978) J. Biol. Chem. 121, 339-356

Blake, C. C. F. & Oatley, S.J. (1977) Nature 268, 115-120

Bourne, Y., Roussel, A., Frey, M., Rougé, P., Fontecilla-Camps, J-C. & Cambillau, C. (1990) Proteins, 7, 365-376

Bourne, Y., Mazurier, J., Legrand, D., Rougé, P., Montreuil, J., Spik, G. & Cambillau, C. (1994) Structure , 2, 209-219

Brünger, A.T. (1992) X-PLOR version 3.1 : A system for crystallography

and NMR. Yale University, New Haven, CT.

Brünger, A.T. (1993) Acta Cryst., D49, 24-36

Casset, F., Hamelryck, T., Loris, R., Brisson, J., Tellier, C., Dao-Thi, M., Wyns, L., Poortmans, F., Pérez, S. & Imberty, A. (1995) J. Biol. Chem., 270, 25619-25628

Chothia, C. & Janin, J. (1981) Proc. Natl. Acad. Sci. USA, 78, 4146-4150

Chrispeels, M. J. & Raikhel, N. V. (1991) Plant Cell 3, 1-9

Ciszak, E., Cody, V. & Luft, R. L. (1992) Proc. Natl. Acad. Sci. USA 89, 6644-6648

Dao-Thi, M.-H., Hamelryck, T. W., Poortmans, F., Voelker, T. A., Chrispeels, M. J. & Wyns, L. (1996) Prot. Struct. Funct. Gen. 24, 134-137

Delbaere, L., Vandonselaar, M., Prasad, L., Quail, W., Pearlstone, J., Carpenter, M., Smillie, L., Nikrad P., Spohr, U. & Lemieux, R. (1990) Can. J. Chem., 68, 1116-1121

Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F. & Sacchettini, J. C. (1995) Biochemistry 34, 4933-4942

Edelman, G.M. & Wang, J.L. (1978) J. Biol. Chem., 253, 3016-3022

Ferguson, R. N., Edelhoch, H., Saroff, H. A. & Robbins, J. (1975) Biochemistry 14, 282-289

Gabriel, C. (1950) Z. Lebensm. Untersuch. Forsch., 90, 191-197

Gallagher, T., Rozwarski, D. A., Ernst, S. R. & Hackert, M. L. (1993) J. Biol. Chem. 230, 516-528

Gegg, C. V., Roberts, D. D., Segel, I. H. & Etzler, M. E. (1992) Biochemistry 31, 6938-6942

Gegg, C. V. & Etzler, M. E. (1994) J. Biol. Chem. 269, 5687-5692

Goossens, A., Geremia, R., Bauw, G., Van Montagu, M. & Angenon, G. (1994) Eur. J. Biochem. 225, 787-795

Hammerström, S., Hammerström, M., Sundblad, G., Arnarp, J. & Lönngren, J. (1982) Proc. Natl. Acad. Sci. USA 79, 1611-1615

Hartweck, L. M., Vogelzang, R. D. & Osborn, T. C. (1991) Plant Physiol. 97, 204-211

Hoffman, L.M. & Donaldson, D.D. (1985) EMBO J. 4, 883-889

Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991) Acta Cryst, A47, 110-119

Jordan, E. T. & Goldstein, I. J. (1994) J. Biol. Chem. 269, 7674-7681

Kraulis, P. J. (1991) J. Appl. Cryst. 24, 946-950

Kella, N.K.D., Roberts, D.D., Shafer, J.A. & Goldstein, I.J. (1984) J. Biol. Chem., 259, 4777-4781

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) J. Appl. Cryst. 26, 283-291

Licastro, F., Davis, L.J. & Morini, M.C. (1993) Int. J. Biochem., 25, 845-852

Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R. & Wyns, L. (1993) Biochemistry 32, 8772-8980

Loris, R., Stas, P. P. G. & Wyns, L. (1994) J. Biol. Chem. 269, 26722-26733

Maliarik, M.J., Roberts, D.D. & Goldstein, I.J. (1987) Arch. Biochem. Biophys., 255, 194-200.

Maliarik, M.J. & Goldstein, I.J. (1988) J. Biol. Chem., 263, 11274-11279.

Maliarik, M., Plessas, N.R., Goldstein, I.J., Musci, G. & Berliner, L.J. (1989) Biochemistry, 28, 912-917.

Mandal, D. K., Nieves, E., Bhattacharyya, L., Orr, G. A., Roboz, J., Yu, Q. & Brewer, C. F. (1994) Eur. J. Biochem. 221, 547-553

McDonald, I. K., Naylor, D. N., Jones, D. T. and Thornton, J. M. (1993) 'HBPLUS', Computer Program, Department of Biochemistry and Molecular Biology, University College London

Miller, S. (1989) Protein Engineering 3, 77-83

Mirkov, T.E., Wahlstrom, J.M., Hagiwara, K., Finardi-Filho, F., Kjemtrup, S. & Chrispeels, M.J. (1994) Plant Mol. Biol., 26, 1103-1113

Moreno, J. & Chrispeels, M. J. (1989) Proc. Natl. Acad. Sci. USA 86, 7885-7889

Murzin, A. G. (1994) Curr. Opin. Struct. Biol. 4, 441-449

Navaza, J. (1994) Acta Cryst. A50, 157-163

Nicholls, A., Sharp, K. A. & Honig, B. (1991) Proteins 11, 281-296

Osborn, T. C., Alexander, D. C., Sun, S. S. M., Cardona, C. & Bliss, F. A. (1988) Science 240, 207-210

Pflugrath, J. W. & Messerschmidt, A. (1989) MADNESS manual of FAST diffractometer. Enraf-Nonius, Delft, The Netherlands

Pusztai, A. & Bardocz, S. (1996) TIGG, 8, 149-165

Rini, J.M., Hardman, K.D., Einspahr, H., Suddath, F.L. & Carver, J.P (1993) J. Biol. Chem. 268, 10126-10132

Roberts, D. D., & Goldstein, I. J. (1982) J. Biol. Chem. 257, 11274-11277

Roberts, D. D. & Goldstein, I. J. (1983a) J. Biol. Chem. 258, 13820-13824

Roberts, D. D. & Goldstein, I. J. (1983b) Chemical taxonomy, Molecular Biology, and Function of Plant Lectins (Goldstein, I. J., and Etzler, M. E., Eds.), 131-141, Alan R. Liss, New York

Romero Andreas, J., Yandell, B.S. & Bliss, F.A. (1986), Theor. Appl. Genet. 72, 123-128

Sembdner, G., Atzorn, R. & Schneider, G. (1994) Plant Mol. Biol., 26, 1459-1481

Shaanan, B., Lis, H. & Sharon, N. (1991) Science, 254, 862-866

Shade, R. E., Schoeder, H. E., Pueyo, J. J., Tabe, L. M., Murdock , L. L., Higgins, T. J. V. & Chrispeels, M. J. (1994) BIO/TECHNOLOGY 12, 793-796

Sturm, A. & Chrispeels, M. J. (1986) Plant Physiol. 80, 320-322

Van Damme, E.J.M., Peumans, W.J., Pusztai, A. & Bardocz, S. (1997). Handbook of plant lectins: properties and biomedical applications. John Wiley & Sons, New York.

Wojtczak, A., Luft, J. & Cody, V. (1992) J. Biol. Chem. 267, 353-357

Yang, D.C.H., Gall, W.E. & Edelman, G.M. (1974) J. Biol. Chem., 249, 7018-7023

Young, N. M., Watson, D. C., Yaguchi, M., Adar, R., Arango, R., Rodriguez-Arango, E., Sharon, N., Blay, P. K. S. & Thibault, P. (1995) J. Biol. Chem. 270, 2563-2570

Chapter 3

Arcelin-5, a truncated lectin from the common bean with insecticidal activity

Reference: Hamelryck, T. W., Poortmans, F., Goossens, A., Angenon, G., Van Montagu, M., Wyns, L., & Loris, R. (1996) Crystal structure of arcelin-5, a lectin-like defense protein from Phaseolus vulgaris. **J. Biol. Chem.**, 271, 32796-32802

3.1 Introduction

In the seeds of the common bean, *Phaseolus vulgaris*, two legume lectins are found, called phytohemagglutinin-L (PHA-L) and phytohemagglutinin-E (PHA-E). The crystal structure of PHA-L is described in the previous chapter. In addition, two proteins which show high sequence identity with these lectins are present in the seeds, namely arcelin and a-amylase inhibitor (αAI). In these so-called lectin-like proteins, respectively one and two loops that are essential for sugar binding are deleted. Therefore, αAI and arcelin do not bind sugars and are not lectins. αAI inhibits α-amylases of mammalian and insect origin, but does not inhibit plant α-amylases. Arcelin is only present in a few wild accessions of the common bean, and exists in six electrophoretic variants. The phytohemagglutinins, αAI and arcelin are defense proteins that protect the bean against various predators (Chrispeels & Raikhel, 1991), although the mechanism behind the toxicity of arcelin is as yet unknown. The four proteins are encoded by four tightly linked genes, and it is likely that these originate from a common ancestral gene through duplication (Hoffman & Donaldson, 1985; Mirkov *et al.*, 1994).

Arcelins protect wild bean strains against the bean bruchid pest *Zabrotes subfasciatus* or Mexican bean weevil (Osborn *et al.*, 1988; Fig. 3.1). The larvae of the Mexican bean weevil burrow into the seeds and are responsible for considerable post harvest losses in the Americas. The presence of arcelin in the bean seeds increases mortality of Mexican bean weevil larvae and prolongs the emergence of the adult beetle. Among the arcelin vari-

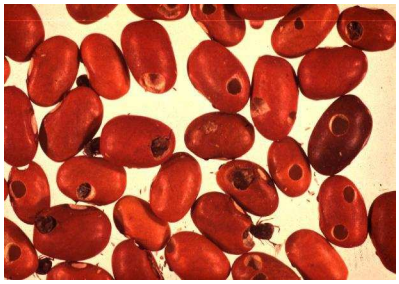


Figure 3.1: Beans infected with *Zabrotes subfasciatus*. The adult beetles are approximately 4 mm long. When the adult beetle leaves the bean, it cuts out a small window in the integument of the seed, which is clearly visible on the picture.

ants, Arc1 and Arc5 appear to be the most promising in conferring insect resistance (Kornegay *et al.*, 1993). In this chapter the crystal structure of arcelin-5 solved at a resolution of 2.7 Å is presented.

Arcelin-5 (Goossens *et al.*, 1994) has been reported to consist of a mixture of two major protein fractions, termed arcelin-5a (Arc5a, 32.2 kD) and arcelin-5b (Arc5b, 31.5 kD), and a third, minor fraction, termed arcelin-5c (Arc5c, 30.8 kD). Arc5b and Arc5a contain one and two glycans, respectively, while Arc5c is not glycosylated. Two different arcelin-5 cDNA sequences were reported (Goossens *et al.*, 1994), called arc5-I and arc5-II. They encode two polypeptides of 240 amino acids (26.8 and 27.0 kD) with a high identity (96.9 %, a difference of 8 residues in the N terminal part of the chain). The arc5-I encoded protein contains three potential glycosylation sites, while the arc5-II encoded protein contains only two. Arc5a and Arc5b are encoded by arc5-I and arc5-II, respectively, while Arc5c could be encoded by arc5-II or by a third copy of the arc5 gene with a much lower rate of expression.

3.2 Materials and methods

3.2.1 Crystallization and data collection

Arcelin-5 was purified from seeds of the wild *Phaseolus vulgaris* accession G02771 by chromatofocusing as described in (Goossens *et al.*, 1994). Protein solutions were made starting from the lyophylised protein. Suitable crystallization conditions were screened using the hanging drop vapor diffusion method : 5 μl aqueous protein solution (varying between 5 and 10 mg/ml) was mixed with 5 μl of various bottom solutions and equilibrated against 0.5 ml of pure bottom solution. Small, needle-like crystals were obtained with a bottom solution consisting of 20 % (w/v) PEG 4000 (Hampton research), 0.2 M $(NH_4)_2SO_4$, Na-acetate buffer pH 4.5, in combination with a 5.0 mg/ml aqueous protein solution, incubated at room temperature. To obtain crystals suitable for data collection, the micro-seeding technique was used. A 10 μl drop containing many needle-like crystals was homogenized by vortexing and a range of dilutions between 10^{-1} and 10^{-8} in undiluted bottom solution was made. One μl of these dilutions was subsequently added to the 10 μl drops with aforementioned composition prior to equilibration. This resulted in a number of thin, plate-shaped

Table 3.1: Data collection statistics.

Resolution (Å)	Number of observed reflections	Number of unique reflections	Completeness (%)	Multiplicity	Rmerge (%)
6.63-10.0	2892	843	98.9	3.4	4.3
5.31-6.63	3835	1102	98.0	3.5	5.3
4.56-5.31	4468	1325	98.4	3.4	5.5
4.05-4.56	5161	1516	98.8	3.4	6.0
3.69-4.05	5424	1683	99.0	3.2	7.3
3.40-3.69	5519	1830	98.9	3.0	9.2
3.18-3.40	5468	1926	97.1	2.8	12.1
2.99-3.18	5522	2093	98.9	2.6	16.1
2.83-2.99	5526	2173	97.1	2.5	24.4
2.70-2.83	5274	2231	94.8	2.4	32.0
Total	49089	16722	95.9	2.9	8.6

crystals. Best results were obtained using the 10^7-fold dilution. One of the thus obtained crystals was transferred to a glass capillary and used for data collection. This was done on an Enraff-Nonius FAST area detector, using Cu Kα radiation generated by a rotating anode X-ray generator, operated at 40kV and 90 mA. Autoindexing was done with the MADNESS software packet (Pflugrath & Messerschmidt, 1989). The crystal was monoclinic and belonged to space group $P2_1$. The data was subdivided in bins of 3° and scaled and merged with the CCP4 programs ROTAVATA and AGROVATA (Collaborative computational project, n. 4, 1994). The relevant data collection details are given in table 3.1.

3.2.2 Structure solution and refinement

The structure of Arc5 was solved using molecular replacement with the program AMORE (Navaza, 1994). The co-ordinates of PHA-L (Hamelryck et al., 1996) were used as a model, after removal of sugar, water and metal atoms. PHA-L is a tetramer consisting of two canonical legume lectin dimers. Since Arc5 has been reported to be a dimer in solution (Goossens *et al.*, 1994) and the identity between Arc5 and PHA-L is high (55 %), molecular replacement was first attempted, however unsuccesfully, with a complete PHA-L canonical dimer. Conversely, two clear solutions were found using the PHA-L monomer as a model. Application of the found solutions to the model and visual inspection showed reasonable crystal packing contacts and the absence of significant sterical clashes.

Since Arc5 shows high sequence identity with members of the legume lectin family, the Arc5 sequence was submitted to the Swiss-Model service (Peitsch, 1996) for automated comparative modeling. This model was used for further refinement. During the refinement the PHA-L structure was often used to clarify pieces of the structure where the density was hard to interpret.

The refinement was done with the program X-PLOR (version 3.1) (Brunger, 1992) on a Cray J916/8-1024 supercomputer. Positional refinement was done with the simulated annealing protocol (SA) (4000 K, Δt=0.01 ps). In

Table 3.2: Overview of the statistics of data collection and refinement.

Space group	$P2_1$
Unit cell dimensions	a=41.30 Å b=94.50 Å c=82.90 Å β=94.97 °
Resolution	2.7 Å
Final R-value (10.0 Å-2.7 Å)	20.6 %
Final Free R value (10.0 Å-2.7 Å)	27.1 %
Number of protein atoms	4554
Number of solvent atoms	0
Number of glycan atoms	76
R.m.s on bond lengths	0.017 Å
R.m.s. on bond angles	2.17 °
R.m.s. on dihedrals	28.22 °
R.m.s. on impropers	2.06 °
Distribution of non-Gly non-Pro ϕ-ψ angles in the Ramachandran plot	82.2 % in core regions 17.4 % in additionally allowed regions 0.4 % in generously allowed regions 0.0 % in disallowed regions

order not to overfit the structure in the early stage of the refinement, the grouped B factor refinement protocol (one B factor for side chain atoms and one B factor for main chain atoms per residue) was used in the first four refinement cycles. After cycle 4, individual B factor refinement was applied since the difference in free R factor between grouped and individual B factor refinement clearly justified this (free R factor=32.0 %, R factor=24.7 % for the grouped B factor refinement, free R factor=31.0 %, R factor=23.6 % for the individual B factor refinement). In accordance with the currently accepted refinement protocols (Kleywegt & Jones, 1995), restrained non-crystallographic symmetry (weight=300.0 kcal mol^{-1} Å^{-2}, σ_{ncs}=1.0 Å^2) was applied between the two monomers in the asymmetric unit. During the refinement SA omit maps were extensively used to clarify parts of the structure with difficult to interprete electron densities. Model visualization and building were done using the program O (version 5.10.2; Jones *et al.*, 1991) on a Silicon Graphics INDIGO^2 XZ workstation. The stereochemistry of the structure was monitored throughout the refinement with the program PROCHECK (Laskowski *et al.*, 1993). Further details about the structure are given in table 3.2.

The co-ordinates of GS4, EcorL and PHA-L, used for comparison with Arc5, are present in the Brookhaven protein database under entries 1LEC, 1LTE and 1FAT, respectively. Hydrogen atoms on the sugar residues, necessary for the determination of the torsion angles, were generated with X-PLOR. Hydrogen bonds were determined with the program HBPLUS (McDonald *et al.*, 1993). The figures were made with the programs MOLSCRIPT (Kraulis, 1991) and O (Jones *et al.*, 1991). The co-ordinates and structure factors of Arc5 are available from the Brookhaven Protein Databank under entries 1IOA and r1IOAsf, respectively.

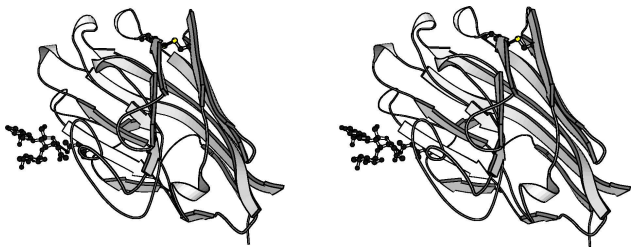


Figure 3.2: An overall view of the Arc5 monomer. The glycan attached to Asn 22 (GlcNAc(β1–4)[Fuc(α1–3)(GlcNAc) and the disulfide bridge between residues 146 and 182 are shown as ball-and-stick models.

3.2.3 Gel filtration

The molecular mass of the native arcelin-5 protein was determined by gel filtration through a Superdex 200 HR (Pharmacia) FPLC column (1.0 x 30 cm) with a flow rate of 0.8 ml/min. Phosphate buffered saline (0.14 M NaCl, 3 mM KCl, 50 mM Na_2HPO_4, 10 mM KH_2PO_4, pH 7.2) or acetate buffer (50 mM CH_3COONa, pH 4.5) were used as the equilibration and elution buffer. As a standard the gel filtration standard (Bio-rad) with molecular weight markers ranging from 1.35 to 670 kDa was used.

3.3 Results and discussion

3.3.1 Overall structure

The R factor of the final structure for all 16722 reflections between 2.7 and 10.0 Å is 20.6 %, the free R factor is 27.1 %.

In the final cycles of the refinement it became clear that the conformation of the residues between Val 36 and Pro 42 differed between the two monomers in the asymmetric unit, due to different crystal contacts. The two regions were manually rebuilt and were not subjected to NCS during the final refinement cycles. This region in the first monomer is the only region in the structure that displays comparatively weak electron density. Nevertheless, all residues could properly be built into the structure. Because of the low resolution (2.7 Å), no waters were added to the structure during the refinement.

The overall structure of Arc5 is similar to the structures of the other solved legume lectin structures, and consists of a flat six strand β-sheet, called the back sheet, packed against a curved seven strand β-sheet, called the front sheet (see Fig. 3.2). Since NCS was used during the refinement, the two Arc5 monomers are virtually identical, with the exception of the region between Val 36 and Pro 42, due to crystal packing contacts. The structure contains 228 of the 240 amino acids of the mature protein : no electron density is observed for the terminal 12 amino acids. This can be due to C-terminal truncation or dynamic disorder of the C-terminal stretch in the crystal.

Interpretable electron density, for both monomers in the asymmetric unit, was present for the glycan attached to Asn 22, which consists of two GlcNAc and one Fuc residue (for further details; see below). A fourth sugar residue displayed only weak electron density and was subsequently

Figure 3.3: The superposition of Arc5 on the PHA-L monomer. The Arc5 backbone is traced with a bold line; the PHA-L backbone is traced with a thin line. The two bound metals of PHA-L are represented as red spheres. The major difference betwe en the legume lectin monomers and Arc5, i.e. the absence of the metal binding loop and the two bound metals in Arc5 (top of the picture; shown in red), is clearly visible.

not built into the model. Arc5 consists of two polypeptides with slightly different amino acid sequences : 8 of the 240 residues differ between the two sequences. Since a glycan attached to Asn 22 could be refined at full occupancy in both monomers present in the asymmetric unit, resulting in low B factors for the sugar atoms, and given the fact that only the polypeptide encoded by the arc5-I gene contains a potential glycosylation site at position 22, we can safely conclude that the crystal mainly consists of Arc5a. Hence, the arc5-I encoded sequence was used to build the structure. The region that is involved in crystal packing contacts (Val 36-Pro 42) contains four residues that differ between the Arc5a and Arc5b sequences (Gly 37, Ser 38, Glu 40 and Leu 41 for Arc5a; Thr 37, Pro 38, Gly 40 and Asp 41 for Arc5b). This is probably the reason why the Arc5 crystal mainly contains Arc5a.

The Arc5 and PHA-L amino acid sequences have an identity of about 55 %. The r.m.s. difference between the positions of the superimposable Cα atoms in Arc5 and the PHA-L monomers is about 1.1 Å. The differences between the two structures are mainly located in loop regions (Asn 111-Asn 116, Lys 78-Ile 82 and Gly 207-Thr 212) and the N-terminal β-strand (see Fig. 3.3). Most notable is the deletion of eight residues in the metal binding loop (Phe 127-Arg 130), as described in detail below.

The Arc5 monomer contains one disulfide bridge, between residues 146 and 182. These two residues link the sixth strand of the flat front sheet and the seventh strand of the curved back sheet together.

3.3.2 The quaternary structure of Arc5

Although Arc5 has been reported to be a dimer in solution (Goossens et al., 1994), the here described crystal structure of Arc5 suggested that it crystallized as a monomer : no symmetric dimer is present in the crystal. The contact surfaces between the possible dimers in the crystal are of comparable size, and all of them resemble typical crystal packing contact surfaces. Gel filtration indicated that arcelin-5 can exist both as a monomer or as a multimer, eluting at positions corresponding with a molecular weight of 25 kDa or 77 kDa, respectively. The sample used for crystallization consisted of purified protein which had first been lyophilised and then resuspended in water. In this sample only the monomer form was present. This was also the case when the lyophilised protein was resuspended in

phosphate buffered saline or in acetate buffer. In a crude extract (partially purified albumins; Goossens *et al.*, 1994) and in a purified fraction which had never been lyophilised before, the protein existed as a mixture of monomers and multimers.

The range of quaternary structures of the legume lectin family is extremely broad, despite the similarity of the monomers : at present three types of tetramers are known (Con A, Becker *et al.*, 1975; peanut agglutinin, Banerjee *et al.*, 1994; and PHA-L/DBL, Chapters 2 and 4) and four types of dimers (*Erythrina corallodendron* lectin, EcorL, Shaanan *et al.*, 1991; GS4, Delbaere *et al.*, 1989; DB58, Chapter 4; and the group of canonical dimers, e.g. lentil lectin, Loris *et al.*, 1993). The monomeric Arc5 structure brings the number of known quaternary structures to eight.

The presence of glycans has been invoked to explain the different quaternary structures of *Griffonia simplicifolia* isolectin IV (GS4, Delbaere *et al.*, 1989) and EcorL (Shaanan *et al.*, 1991). In any case, the positions of the three potential glycosylation sites do not exclude the possibility of canonical dimer formation in Arc5.

Different arcelins have been reported to have different quaternary structures : arcelin-1d and arcelin-2 have been reported to be dimers, while arcelin-1t, arcelin-3 and arcelin-4 have been reported to be tetramers (Hartweck *et al.*, 1991). The two canonical lectin dimers in the PHA-L structure pack together via intercalation of the side chains of the two top β-strands of the flat back sheet. This intercalation is possible because the side chains of the residues involved are small (one central Ile and two flanking Ser residues). In Arc5, PHA-L like packing would be impossible because one of the two Ser residues present in the PHA-L dimer-dimer interface (Ser 186) has been replaced by Lys 181. Strikingly, all the known sequences of the arcelins reported to be dimers possess this Lys residue (arcelin-5, -2 and -1, assuming that the reported sequence of arcelin-1 in Osborn *et al.*, 1988 is the sequence of arcelin-1d), while in the sequence of arcelin-4, which has been reported to be a tetramer, this Lys residue has been replaced by an Ala residue. In addition, the arcelin-1 and -2 sequences also possess a His residue in the position of the Ile 188 residue in PHA-L.

3.3.3 The glycan attached to Asn 22

The arc5-I encoded protein contains three potential N-glycosylation sites (Asn 22, Asn 70 and Asn 79), while the arc5-II encoded one contains only two (Asn 70 and Asn 79). Electrospray mass spectrometry indicated that Arc5-a contains two glycans, Arc5-b contains one, and Arc5-c contains none (Goossens *et al.*, 1994).

Interpretable density was only observed for the glycan attached to Asn 22 (see Fig. 3.4). Asn 22 is located in the loop that connects the first and second β-strand of the front sheet. As mentioned before, the fact that the glycan can be refined with full occupancy indicates that mainly the arc5-I encoded protein is present in the crystal, since the arc5-II encoded sequence does not possess this glycosylation site. Since Arc5-a contains two glycans, the other glycan has to be invisible due to its inherent flexibility. No electron density whatsoever is found for the two other potential glycosylation sites, so it is impossible to say where the other glycan is attached.

Of the glycans attached to Asn 22, only the three core residues were suf-

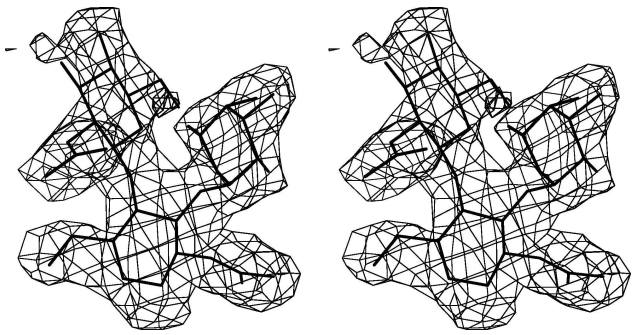


Figure 3.4: The $2F_o$-F_c electron density around the glycan (GlcNAc(β1-4)[Fuc(α1-3) (GlcNAc) bound to Asn 22. The electron density was visualized at the 0.8 σ level.

Table 3.3: The Φ and Ψ torsion angles (Φ/Ψ) of the GlcNAc(β1–4)[Fuc(α1–3)]GlcNAc core of the glycan moieties of GS4, EcorL, Arc5 and bromelain. The latter have been determined by NMR measurements on the glycan in solution.

	GS4	EcorL	Arc5	Bromelain
Fuc(α1-3)GlcNAc	30/18	45/19	48/10	45/30
GlcNAc(β1-4)GlcNAc	51/7	47/11	34/3	50/10

ficiently visible. The electron density of these three visible sugar residues (GlcNAc(β1–4)[Fuc(α1–3)]GlcNAc) indicate that the glycan is of the fucosylated, complex type. The Fuc residue bound (α1–3) to the N-linked GlcNAc residue is exclusively found in plant N-glycans. The same sugar type is also present in the GS4 (Delbaere *et al.*, 1989) and EcorL (Shaanan *et al.*, 1991) structures. In the former structure only the same three core residues as in the Arc5 structure are visible, while in the latter structure a further four residues could be built in (Man(α1–6)[Man(α1–3)][Xyl(β1–2)]Man attached by a (β1–4) bond to the GlcNAc residue of the core), because the flexibility of the glycan was hampered by fortunate crystal packing interactions. In Arc5, the two GlcNAc residues of the glycan make four hydrogen bonds with three protein residues (Asn 22, Arg 43 and Lys 101). No protein-sugar interactions are present for the Fuc residue. The glycan makes five internal hydrogen bonds. The three sugar residues are in the expected $^4C_1$ conformation. Manual superposition of the GlcNAc(β1–4)[Fuc(α1–3)]GlcNAc core which the glycans from GS4, EcorL and Arc5 have in common, revealed that the conformations of the GS4 and the EcoRL cores are virtually identical, while the conformation of the Arc5 core differs somewhat from them.

Bouwstra *et al.* (1990) studied the conformational characteristics in solution of the glycan attached to bromelain, a proteolytic enzyme from pineapple stem, with ^{1}H- and ^{13}C-NMR. The studied glycan resembles the EcorL glycan : only the (α1–3) bound Man residue is absent. The Φ and Ψ torsion angles for both glycosydic bonds of the three residue core which the glycans from Arc5, GS4, EcorL and bromelain have in common, are shown in table 3.3.

The Φ and Ψ torsion angles are defined as (C1-O-Cx-Hx) and (H1-

C1-O-Cx), respectively, where x equals 3 for Fuc(α1–3)GlcNAc and 4 for GlcNAc(β1–4)GlcNAc. The difference between the Arc5 glycan and the other glycans lays mainly in the conformation of the Fuc(α1–3)GlcNAc glycosidic bond.

The fact that only the core GlcNAc(β1–4)[Fuc(α1–3)]GlcNAc moiety is observed in both Arc5 and GS4 may be due to the inherent rigidity of this moiety. Imberty *et al.* (1990, 1991) studied the conformations of the different disaccharides present in glycans using molecular modelling and energy calculations. For each disaccharide a conformational energy map in function of the j and y values was calculated. For the GlcNAc(β1–4)GlcNAc disaccharide six low energy conformations were found (A1-A6) (Imberty *et al.*, 1990). Plotting the Φ/Ψ pairs of Arc5, GS4, EcorL and bromelain on the energy map reveals that the four conformations cluster closely around the reported A1 energy minimum. For the Fuc(α1–3)GlcNAc disaccharide only three energy minima were found (Imberty *et al.*, 1991). In general, this disaccharide is conformationally more restricted then the GlcNAc(β1–4)GlcNAc disaccharide. Again the Φ/Ψ pairs of the four glycans cluster around a reported energy minimum (L1).

As a conclusion, it can be said that the results about the GlcNAc(β1–4)[Fuc(α1–3)]GlcNAc conformation obtained from NMR measurements in solution, molecular modelling and X-ray crystallography are in close agreement with each other.

3.3.4 The truncated metal binding site

One of the common features of the legume lectin family is the presence of a metal binding site, in which two metal ions are bound (one Ca^{2+} ion and one transition metal ion, usually treated in the X-ray structures as Mn^{2+}). Each metal ion interacts with four residues (Glu 122, Asp 124, Leu 126, Asn 128, Asp 132 and His 137 in PHA-L) and two water molecules (see Fig. 3.5). The side chains of two Asp residues bridge the two bound metal ions (Asp 124 and Asp 132 in PHA-L). The bound Ca^{2+} ion interacts via a water molecule with the side chain oxygen and the main chain carbonyl group of a conserved Asp residue, and thus stabilizes a *cis*-peptide bond between this Asp residue and an Ala residue (Ala 85-Asp 86 in PHA-L). Removing the bound metals results in a *cis* to trans isomerization of the peptide bond between the Ala and the Asp residue in Con A (Bouckaert *et al.*, 1995). The sugar binding capabilities of the legume lectins depend on the presence of this *cis*-peptide bond, and thus on the presence of the bound metals.

In Arc5, five of the six conserved metal ligands are missing or not appropriate for metal ligation. One Ca^{2+} ligating Asn residue (Asn 128 in PHA-L) and one bridging Asp residue (Asp 132 in PHA-L) are both situated in a deleted region in Arc5, and are thus absent. The two Mn^{2+} ligating His and Glu residues (His 137 and Glu 122 in PHA-L) have been replaced by Arg 130 and Val 123 in Arc5, respectively. The second bridging Asp residue (Asp 124 in PHA-L) has been replaced by Asn 125 in Arc5, thereby removing the stabilizing negative charge. Indeed , no evidence for bound metal ions can be found in the electron density.

In addition, the highly conserved Asp residue (Asp 86 in PHA-L) involved in the *cis*-peptide bond is replaced by Tyr 85 in Arc5. Therefore it is very surprising that a *cis*-peptide bond is still present between the residues Ala 84 and Tyr 85, despite the absence of the two bound metal ions. The *cis*-

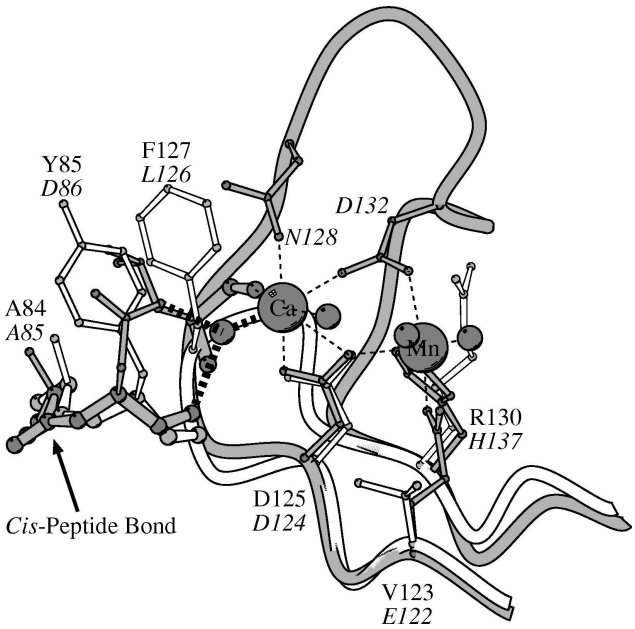


Figure 3.5: The metal binding site of PHA-L superimposed on the truncated metal binding site of Arc5. The backbone of Arc5 is outlined in white, the backbone of PHA-L in gray. The amino acids belonging to PHA-L are labelled in italic Only the side chains of the relevant amino acids involved in metal ligation or the *cis*-peptide bond are shown as a ball-and-stick representation. The interactions between the metal ions and the amino acids in PHA-L are shown as dotted lines. The interaction between the Ca^{2+} ion and Asp 86 via a water molecule is shown in bold. The figure clearly shows the stacking interaction between Tyr 85 and Phe 127 residue in Arc5.

peptide bond is stabilized by the stacking of the Tyr 85 residue with the nearby Phe 127 residue, and two main chain-main chain hydrogen bonds : between Thr 205 O and Ala 84 NH (2.87 Å in monomer 1, 2.78 Å in monomer 2) and between Gly 207 NH and Ala 84 O (3.21 Å in monomer 1, 3.12 Å in monomer 2), the former being clearly the stronger one. The two phenyl rings from Tyr 85 and Phe 127 are at an angle of approximately 50°. The Phe 127 equivalent residues in the legume lectins are always involved in hydrophobic interactions with the bound sugar.

Fig. 3.6 shows an F_o-F_c, Φ_c SA omit map calculated with the two residues involved in the *cis*-peptide bond, their two neighboring residues (Ser 83 and Gly 86) and the interacting amino acids (Thr 205-Gly 207 and Phe 127) deleted. The electron density around the Tyr 85 side chain and the Ala 84-Tyr 85 main chain atoms is clearly visible and unambiguous, proving the presence of this *cis*-peptide bond.

Cis-peptide bonds are quite rare : in Stewart *et al.* (1990) only 0.36 % of all the peptide bonds in the set of proteins studied was found to be *cis*. Interestingly, Tyr is the residue most often involved in X-Pro *cis*-peptide bonds : 19 of the 76 (25.0 %) Tyr-Pro bonds studied in are *cis*, followed by Ser-Pro (11.0 %) and Phe-Pro (9.6 %). Non X-Pro *cis*-peptide bonds are even rarer : only 0.05 % was found to be *cis*. These rare non X-Pro *cis*-peptide bonds are often involved in catalysis or play some other important functional role (Herzberg & Moult, 1991). As far as we know, this is the

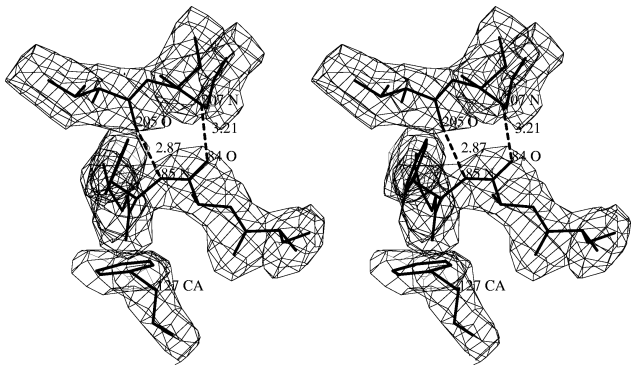


Figure 3.6: Stereo figure of the F_o-F_c, Φ_c SA omit map around the two residues involved in the *cis*-peptide bond, Ala 84 and Tyr 85, and the three residues that s tabilize the *cis*-peptide bond through stacking interactions with the Tyr side chain (Phe 127) or hydrogen bonds (Thr 205 and Gly 207). The map is visualized at the 1.5 σ level and clearly shows the presence of the *cis* conformation.

first report of an Ala-Tyr *cis*-peptide bond.

At present, the mechanism behind the toxicity of the arcelins towards insects remains obscure. The fact that the Ala 84-Tyr 85 peptide bond is still in the *cis* position, despite the absence of a metal binding site and the low frequency of *cis*-peptide bonds in proteins, might point to a possible involvement of these residues in the mode of action of this protein. It is also striking that in all five known arcelin sequences (Arc1, Arc2, Arc4 and Arc5 from *Phaseolus vulgaris* and a fifth arcelin-like protein from *Phaseolus acutifolius*, Mirkov *et al.*, 1994) the Ala residue followed by an aromatic residue (Phe in the case of Arc4, Tyr in the four other cases) is found.

Concanavalin B (Con B), a seed protein from *Canavalia ensiformis*, and hevamine, a chitinase from Hevea brasiliensis, show an interesting and striking analogy to the situation of Arc5 and the legume lectins. Despite a high sequence identity (40 %), hevamine is active as a chitinase while Con B appears to lack any enzymatic activity, due to the replacement of a catalytic Glu residue by a Gln residue and the partial deletion of the substrate binding region (Hennig *et al.*, 1995). Yet, Con B still possesses two non-proline *cis*-peptide bonds (Ser 34-Phe 35 and Trp 265-Asn 266) that are a conserved feature of this family of chitinases. In hevamine, the residues equivalent to Phe 35 and Trp 265 in Con B (Phe 32 and Trp 255) are part of a hydrophobic cluster that is involved in substrate binding. It has therefore been suggested that these *cis*-peptide bonds play a role in some unknown function of the Con B protein (Hennig *et al.*, 1995).

3.4 Conclusions

The structure of arcelin-5, a plant defense protein from *Phaseolus vulgaris* related to the legume lectin family, is solved at a resolution of 2.7 Å. Arc5 crystallizes as a monomer whose structure is similar to the structure of the legume lectin monomers, but with the metal binding loop deleted. One of the most striking features of this structure is the conservation of a *cis*-peptide bond that is found in all the legume lectins, despite the absence

of the metal binding site that normally stabilizes this bond in the legume lectins. While this *cis*-peptide bond is always between an Ala and an Asp residue in the true legume lectins, the Asp residue has been replaced by a Tyr residue in Arc5. Inspection of the five available arcelin sequences revealed that they all possess an Ala residue followed by an aromatic residue (Tyr or Phe). Given the fact that only 0.05 % of all non X-Pro peptide bonds are *cis* (Stewart *et al.*, 1990) and that these *cis*-peptide bonds often play some functional role (Herzberg & Moult, 1991), the likely conservation of this feature suggests that it could be involved in the (unknown) mode of action of the protein.

3.5 Notes

Refinement of the arcelin-5 structure was begun when there still was considerable debate about the usefulness and applicability of the R_{free} factor. Therefore, the structure was refined using all reflections. After refinement, an a posteriori R_{free} factor was calculated by performing a simulated annealing refinement with 10 % of the reflections omitted and calculating the R factor.

The recently published structure of arcelin-1 at 1.9 Å resolution (Mourey *et al.*, 1998) reveals that the Ala-Tyr *cis*-peptide bond is also present in this arcelin variant. The Ala-Tyr/Phe *cis*-peptide bond indeed seems to be a conserved feature of the arcelins. It remains to be proven however whether it is also directly involved in the insecticidal activity of the arcelins.

3.6 References

Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Surolia, A., and Vijayan, M. (1994) Proc. Natl. Acad. Sci. USA 91, 227-231

Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A., & Edelman, G. M. (1975) J. Biol. Chem. 250, 1513-1524

Bouckaert, J., Loris, R., Poortmans, F., Wyns, L. (1995) Proteins Struct. Func. Gen. 23, 510-524

Bouwstra, J. B., Spoelstra, E. C., De Waard, P., Leeflang, B., Kamerling, J. P. & Vliegenthart, J. F. G. (1990) Eur. J. Biochem. 190, 113-122

Brünger, A. T. (1992) X-PLOR version 3.1 : A system for crystallography and NMR, Yale University, New Haven, CT

Chrispeels, M. J. & Raikhel, N. V. (1991) Plant Cell 3, 1-9

Collaborative computational project, n. 4, SERC Daresbury laboratory (1994) Acta Cryst. D50, 760-763

Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Nikrad, P. V., Pearlstone, J. R., Carpenter, M. R., Smillie, L. B., Spohr, U. & Lemieux, R. U. (1989) Transactions ACA 25, 65-76

Dessen, A., Gupta, D., Sabesan, S., Brewer, F. & Sacchettini, J. C. (1995) Biochemistry 34, 4933-4942

Goossens, A., Geremia, R., Bauw, G., Van Montagu, M. & Angenon, G. (1994) Eur. J. Biochem. 225, 787-795

Hamelryck, T. W., Dao-Thi, M., Poortmans, F., Chrispeels, M. J., Wyns, L. & Loris, R. (1996) J. Biol. Chem.271, 20479-20485

Hartweck, L., Vogelzang, R. & Osborn, T. (1991) Plant Physiol. 97, 204-211

Hennig, M., Jansonius, J. N., Terwisscha van Scheltinga, A. C., Dijkstra, B. W., & Schlesier, B. (1995) J. Mol. Biol. 254, 237-246

Herzberg, O. & Moult, J. (1991) Prot. Struct. Func. Gen. 11, 223-229

Hoffman, L.M. & Donaldson, D.D. (1985) EMBO J., 4, 883-889

Imberty, A., Delage, M., Bourne, Y., Cambillau, C. & Pérez, S. (1991) Glycoconjugate J. 8, 456-483

Imberty, A., Gerber, S., Tran, V. & Pérez, S. (1990) Glycoconjugate J. 7, 27-54

Jones, T. A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991) Acta Cryst. A47, 110-119

Kleywegt, G. ,J. & Jones, T. ,A. (1995) Structure 3, 535-540

Kornegay, J., Cardona, C. & Posso, C. (1993) Crop Sci. 33, 589-594

Kraulis, P. J. (1991) J. Appl. Cryst. 24, 946-950

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) J. Appl. Cryst. 26, 283-291

Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R. & Wyns, L. (1993) Biochemistry 32, 8772-8980

McDonald, I. K., Naylor, D. N., Jones, D. T. & Thornton, J. M. (1993) 'HBPLUS', Computer Program, Department of Biochemistry and Molecular Biology, University College London

Mirkov, T. E., Wahlstrom, J. M., Hagiwara, K., Finardi-Filho, F., Kjemstrup, S. & Chrispeels, M. J. (1994) Plant Mol. Biol. 26, 1103-1113

Mourey, L., Pédelacq, J.D., Birck, C., Fabre, C., Rougé, P. & Samana, J.P. (1998) Crystal structure of the arcelin-1 dimer from *Phaseolus vulgaris* at 1.9-Å resolution. J Biol Chem, 273, 12914-12922.

Navaza, J. (1994) Acta Cryst. A50, 157-163

Osborn, T. C., Alexander, D. C., Sun, S. S. M., Cardona, C. & Bliss, F. A. (1988) Science 240, 207-210

Peitsch, M. C. (1996) Biochem. Soc. Trans. 24, 274-279

Pflugrath, J. W. & Messerschmidt, A. (1989) MADNESS manual of FAST diffractometer, Enraff Nonius, Delft, The Netherlands

Shaanan, B., Lis, H. & Sharon, N. (1991) Science 254, 862-865

Chapter 4

DBL and DB58 from *Dolichos biflorus*

Reference: 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, 1161-117

4.1 Introduction

The *Dolichos biflorus* seed lectin (DBL) is one of at least four blood group A+H substance binding lectins present in this plant (Etzler, 1996). It is a tetrameric glycoprotein with a molecular mass of 110 kD that is composed of two types of subunits, designated subunit I and II (Carter & Etzler, 1975). Subunit II (241 AA) is post-translationally formed from subunit I (253 AA) by the removal of twelve amino acids from its C-terminus (Roberts, *et al.*, 1982; Young, *et al.*, 1995).

DBL only agglutinates epitopes with terminal nonreducing GalNAc residues and is unique among the GalNAc binding legume lectins in its extreme preference for GalNAc over Gal (Etzler & Kabat, 1970; Hammarström, et al., 1977). Me-α-D-GalNAc is a two-fold better inhibitor of binding in a solid phase assay than GalNAc (Etzler, 1994b) and binds to DBL with an association constant of 4.2 10^3 M^{-1} (Etzler, *et al.*, 1981). Me-α-D-GalNAc, GalNAc(α1–3)Gal and GalNAc(α1–3)Gal(β1–3)GlcNAc are equally active as inhibitors in precipitation assays (Etzler & Kabat, 1970). However, a GalNAc(α1–3)[Fuc(α1–2)]Gal containing pentasaccharide is a slightly better inhibitor, presumably due to the presence of the (α1–2) linked fucose residue. Subsequent studies showed that the Forssman pentasaccharide GalNAc(α1–3)GalNAc(β1–3)Gal(α1–4)Gal(β1–4)Glc and the Forssman disaccharide GalNAc(α1–3)GalNAc are respectively 60- and 40-fold better inhibitors than GalNAc (Baker *et al.*, 1983). The Forssman disaccharide is present in tumor-associated Forssman glycolipid (Hakomori, 1984) and in lipopolysaccharides from certain bacteria, e.g., *Vibrio mimicus* (Landersjö *et al.*, 1998).

Some legume lectins possess a hydrophobic binding site that binds adenine and adenine derived plant hormones, i.e., cytokinins (Roberts & Goldstein, 1983). Binding of adenine does not interfere with carbohydrate

binding (Gegg *et al.*, 1992) and has mainly been described for tetrameric legume lectins, including PHA-L and PHA-E from common bean, soybean agglutinin (SBA) and lima bean lectin (Roberts & Goldstein, 1983; Maliarik & Goldstein, 1988; Maliarik *et al.*, 1989), hog peanut lectin (Maliarik *et al.*, 1987) and *Dolichos biflorus* seed lectin (Gegg *et al.*, 1992). The dimeric stem and leaf lectin from *Dolichos biflorus* (DB58) also binds adenine (Gegg *et al.*, 1992). Binding studies indicate that the mentioned lectins bind adenine and cytokinins with an affinity (K_a=10^5-10^6 M^{-1}) that is two to three orders of magnitude higher than their typical affinity for monosaccharides. The stoichiometry of binding is two adenine molecules per tetramer or one per dimer (Gegg *et al.*, 1992). The preservation of this binding site among legume lectins with different carbohydrate specificities isolated from different plant species suggests that it might be of crucial importance for understanding the role of the legume lectins *in vivo*.

In this chapter, the structures of DBL in complex with the blood group A trisaccharide (GalNAc(α1–3)[Fuc(α1–2)]Gal) and in complex with the Forssman disaccharide (GalNAc(α1–3)GalNAc) are presented, as well as the complex of DBL with adenine which shows that there is an intricate relationship between quaternary structure and adenine binding. In addition, the structure of the adenine binding stem and leaf lectin DB58 (Etzler, 1994a) from the same plant is described. This heterodimeric lectin adopts a novel quaternary structure, related to the structure of the DBL tetramer, in which the hydrophobic binding site is preserved.

4.2 Materials and methods

4.2.1 Preparation of the L127F seed lectin mutant

Site-directed mutagenesis was performed using a two-step PCR method (Landt *et al.*, 1990), employing the oligonucleotide 5'-GCAGTCGAGTTCGACACGTTCTCCAACAGCGGCTGG-3' as the mutagenic primer and using seed lectin cDNA (Schnell & Etzler, 1987) as template. Following sequence confirmation, a Kpn1/BamH1 fragment of the PCR product, containing the Leu127Phe mutation, was cloned into a Kpn1/BamH1 digested pTrc99A based expression vector containing the coding sequence for methionine plus the entire mature seed lectin sequence (Chao et al., 1994). The mutant protein was expressed in the protease deficient SG21173 *E. coli* strain by induction with 0.4 mM IPTG for 3 hours at 37°C. The mutant protein was present in the soluble fraction obtained after sonication and centrifugation of the *E. coli*. This protein was purified by affinity chromatography on hog blood group A+H-Sepharose, followed by specific elution with 0.01 M GalNAc as previously described (Chao *et al.*, 1994).

4.2.2 Binding assays

A solid phase assay (Etzler, 1994b), inhibition of binding of iodinated lectin to bind to hog blood group A+H-Sepharose, was employed for the determination of carbohydrate binding. Various amounts of inhibitors (Gal and GalNAc) were mixed with 72.5 ng of Leu127Phe mutant recombinant seed lectin or native recombinant seed lectin in a final volume of 200 μl of 2% hog blood group A+H-Sepharose. After incubation at room temperature overnight, binding was measured as previously described (Etzler, 1994b).

4.2.3 Crystallization

All DBL complexes were crystallized using recombinant DBL. DBL subunit I was cloned in *E. coli* and purified as previously described (Chao et al., 1994). Crystallisation of the complex of DBL with the blood group A trisaccharide (space group C2), the complex of DBL with adenine (space group $P2_12_12_1$) and uncomplexed DB58 (space group $P2_12_12_1$) is described elsewhere (Dao-Thi *et al.*, 1998). The crystal of DBL in complex with the Forssman disaccharide was made by soaking an uncomplexed $I4_122$ DBL crystal (Dao-Thi *et al.*, 1998) with a 5 mM sugar/3 mM adenine solution. The Forssman disaccharide containing an additional group attached to the anomeric oxygen of the reducing GalNAc (GalNAc(α1–3)GalNAcβ1–$OCH_2CH_2CH_2NH_2$) was purchased from Syntesome, München.

4.2.4 Data collection

Data collection for the DBL sugar complexes was done with a MAR image plate. Cu Kα X-rays were generated with a RIGAKU rotating anode X-ray generator operating at 40 kV, 100 mA. The crystals of the Forssman disaccharide complex diffract anisotropically to 2.4 Å in the c* direction and to 2.8 Å perpendicular to c^*. Refinement was done using data to 2.6 Å. The crystals of the blood group A trisaccharide complex diffract to 2.8 Å.

Data collection for the complex of DBL with adenine (2.65 Å) and for uncomplexed DB58 (3.3 Å) was done with a MAR image plate at the Daresbury synchrotron facility, UK, operating at a wavelength of 1.488 Å.

In all cases, the datasets were indexed, scaled and merged using the programs DENZO and SCALEPACK (Otwinowski & Minor, 1997). Intensities were transformed to structure factors with the CCP4 (Collaborative computational project, SERC Daresbury laboratory, 1994) program TRUNCATE. Details of data set quality are shown in table I.

4.2.5 Molecular replacement, model building and refinement

The phase problem was solved for all structures with the molecular replacement technique, using the CCP4 (Collaborative computational project, SERC Daresbury laboratory, 1994) program AMORE (Navaza, 1994). The PHA-L co-ordinates (Hamelryck *et al.*, 1996) were used for molecular replacement of the DBL-adenine complex. The final model of the DBL-adenine complex was used for molecular replacement of the other three structures (the DBL-Forssman disaccharide complex, the DBL-blood group A trisaccharide complex and uncomplexed DB58).

Molecular replacement of the DB58 dataset was first attempted with a canonical dimer (subunits A and B of DBL, without the helices, see Fig. 4.1). Two solutions were readily found (CC=43.1 %), and inspection of the crystal packing revealed that in the crystal a DBL-like tetramer is formed. However, inspection of an electron density map calculated from this model revealed density for a third non-canonical dimer. A subsequent molecular replacement attempt with an alternative dimer (subunits A and C of DBL, see Fig. 4.1) yielded three clear solutions and a considerable better correlation coefficient (CC=61.7 %).

Positional refinement was done using the simulated annealing, torsion angle refinement and conjugate gradient least squares refinement pro-

Table 4.1: Statistics of crystallographic data and quality of the structures.

	DBL Adenine	DB58 Uncomplexed	DBL Trisaccharide	DBL Forssman
Space group	$P2_12_12_1$	$P2_12_12_1$	C2	$I4_122$
Unit cell parameters a, b, c (Å) α, β, γ(°)	 81.38, 116.14, 224.62 90.00, 90.00, 90.00	 101.39, 130.95, 138.23 90.00, 90.00, 90.00	 96.73, 108.00, 80.96 90.00, 124.67, 90.00	 79.05, 79.05, 260.11 90.00, 90.00, 90.00
Contents of asymmetric unit	2 tetramers 8 adenines	3 dimers	2 monomers 2 GalNAc	1 monomer 1 Forssman
Resolution (Å)	20.0-2.65	20.0-3.3	15.0-2.8	20.0-2.6
I/σI	12.5	9.1	10.7	15.2
I/σI in last shell	2.5, 2.67-2.65 Å	2.7, 3.33-3.3 Å	3.1; 2.82-2.80 Å	2.3; 2.62-2.60 Å
R merge (%)	11.6	13.5	11.1	14.6
Completeness (%)	92.4	80.6	85.2	99.9
Number of reflections measured	162686	47794	39403	91892
Number of unique reflections	57851	22745	14331	13175
R/Rfree (%)	21.9/24.6	22.7/26.6	19.1/21.9	19.4/23.4
R.m.s.d. ideal bond lengths (Å)	0.008	0.011	0.008	0.008
R.m.s.d. impropers (°)	0.67	0.75	0.73	0.63
R.m.s.d. dihedrals (°)	28.4	28.6	28.9	28.7
R.m.s.d. ideal bond angles (°)	1.46	1.58	1.53	1
Ramachandran plot quality (%): Most favored Additionally allowed Generously allowed Forbidden	 89.5 10.1 0.4 0.0	 83.1 16.9 0.0 0.0	 83.1 16.2 0.7 0.0	 84.81 15.2 0.0 0.0

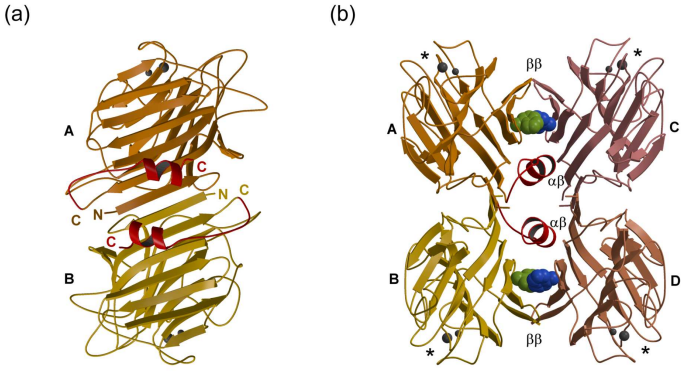


Figure 4.1: (Left) The canonical lectin dimer present in the DBL tetramer. The two subunits are shown in different colors. The sandwiched α-helices (Leu 244-Asn 251), the linker region (Asp 236-Asp 243) and the two C-terminal residues (Val 252 and Leu 253) are shown in red. The part shown in red is only present for the intact subunits. The C-terminal ends of the truncated and the intact subunits are indicated in yellow and red, respectively. The Ca^{2+} and Mn^{2+} ions are shown as large and small gray spheres, respectively. (Right) The overall structure of DBL in complex with adenine. Each subunit is shown in a different color. The four observed adenine molecules are shown as space filling models in green and blue. The two types of dimer-dimer interfaces (the $\beta\beta$-interface and the $\alpha\beta$-interface) are indicated. Dimer AB is shown in the same colors as in Fig. 1a. Metal ions as in Fig. 1a. The locations of the four sugar binding sites are indicated with an asterisk.

tocols of the program X-PLOR (Brünger, 1992). Throughout the refinement a real space bulk solvent model was used. In addition, an overall anisotropic B factor was applied to F_{obs} for both carbohydrate complexes. This improved the quality of the electron maps dramatically in the case of the Forssman disaccharide complex. Non-crystallographic symmetry restraints (σ_b=0.5 Å^2, weight=600 kcal mol^{-1} Å^{-2}) were applied between the subunits in the asymmetric unit for the blood group A trisaccharide complex, the DBL-adenine complex and uncomplexed DB58. The Forssman disaccharide complex only contains one subunit in the asymmetric unit. In order not to overfit the data, only two B factors (one for side chain atoms and one for main chain atoms) per residue were refined for the latter structure. For both sugar complexes, the complete tetramer is generated by the crystal symmetry. The quality parameters of the structures are shown in Table I at the end of this chapter.

Visualisation and model building was done with the program O (version 6.2) (Jones *et al.*, 1991), starting from a modeled DBL subunit (Imberty *et al.*, 1994) for the DBL-adenine complex, and from the resulting DBL coordinates for the other three structures. The amino acid sequences were derived from the cDNA and genomic DNA sequences of DBL and DB58 (Schnell & Etzler, 1987; Schnell & Etzler, 1988; Harada *et al.*, 1990). The DB58 structure was built using an electron density map which was averaged over all subunits in the asymmetric unit with the program DM (Cowtan & Main, 1993). Weak density is present for regions Asn 12-Ser 13, Ser

102-Gly 103 and Asp 131 for the DB58 structure. For the blood group A trisaccharide structure, weak density was present for Ser 79 and Lys 80 in all subunits, except subunits B and H. Consequently, these regions were not built in these structures. Figures were made with the programs BOBSCRIPT (Kraulis, 1991; Esnouf, 1997) and RASTER3D (Merrit & Murphy, 1994). The quality of the structure was analysed with PROCHECK (Laskowski *et al.*, 1993). Accessible surface areas of individual amino acid residues were calculated with NACCESS (Williams *et al.*, 1994). Accessible surface areas of monomer, dimers and tetramers were calculated with the program MSROLL (Connolly, 1993), using a probe size of 1.4 Å.

Torsion angles for the Forssman disaccharide (GalNAc(α1–3)GalNAc) are defined as Ψ=C1-O1-C3'-C4' and Θ=O5-C1-O1-C3'. PDB entries 1AX0 and 1LED were used for comparison of EcorL and GS4 with DBL, respectively.

Coordinates and structure factors have been submitted to the Protein Data Bank (1LU1 and r1LU1sf for the DBL-Forssman disaccharide complex; 1LU2 and r1LU2sf for the DBL-blood group A trisaccharide complex; 1BJQ and r1BJQsf for the DBL-adenine complex; 1LUL/r1LULsf for DB58).

4.3 Results and discussion

4.3.1 Overall structure of DBL

The overall structure of DBL in complex with adenine is shown in Fig. 4.1b. The DBL heterotetramer consists of two intact subunits (253 AA, subunits A and B in Fig. 4.1) and two subunits with a post-translationally truncated C-terminal end (241 AA, subunits C and D in Fig. 4.1) (Schnell & Etzler, 1987; Young *et al.*, 1995). The C-terminal residues of the intact subunits are visible in the electron density, while the last six C-terminal residues of the truncated subunits are not visible.

This quaternary structure was first described for PHA-L (Hamelryck et al., 1996) and is also adopted by SBA (Dessen *et al.*, 1995) and *Vicia villosa* isolectin B4 (Osinaga *et al.*, 1997). Like DBL, the SBA tetramer also consists of C-terminally truncated subunits and intact subunits (Mandal *et al.*, 1994). This tetramer type consists of two canonical legume lectin dimers (dimers AB and CD in Fig. 4.1b) that pack in a parallel fashion against each other. A first interface between the two canonical dimers is present at the outer ends of the tetramer (Fig 1b). This interface, which will be called here the $\beta\beta$-interface, consists of two β-strands that pack together by the zipper-like intercalation of their side chains (Ser 187, Ile 189, Ser 191). The $\beta\beta$-interface diminishes the accessible surface area of the two monomers involved by 1400 Å^2. Essentially the same interface, involving the same Ser-X-Ile-X-Ser motif, is present in SBA and PHA-L (Hamelryck *et al.*, 1996).

However, a second, previously undescribed dimer-dimer interface was observed. The above described mode of association creates a large channel running through the center of the tetramer. In the case of SBA, two stretches of uninterpretable electron density were reported to be present in this channel, presumably due to the C-terminal regions of two of the four subunits (Dessen *et al.*, 1995), while in the case of PHA-L no density is seen (Hamelryck *et al.*, 1996). The corresponding densities in the channel of the DBL tetramer could be interpreted as two α-helices (Leu 244-Asn

251), each formed by the C-terminal part of an intact subunit (Fig. 4.1a and 4.1b). Clear density is also present for the linker region between Asp 235 and Leu 244 and the two C-terminal residues (Val 252 and Leu 253) following the α-helix. The presence of the α-helices in the central channel breaks the apparent 222 symmetry of the tetramer, and changes its point group to 2. The two α-helices are sandwiched between the β-sheets of two pairs of facing monomers (monomer pairs AC and BD in Fig 1b) and form an important stabilization of the tetramer. We will refer to this novel second interface as the $\alpha\beta$-interface. A comparable architecture has recently been observed in a thiolase from yeast (Mathieu *et al.*, 1997), but in that case the sandwiched α-helices belong to an interdomain interface. The unique architecture of the DBL tetramer explains why two truncated and two intact subunits are present per tetramer : the two subunits that have their C-terminal region buried in the central channel remain intact, while the C-terminal regions of the two remaining subunits are proteolitically processed *in vivo* (Schnell & Etzler, 1987; Young et al., 1995). There are some indications that C-terminal processing of half of the subunits is necessary before the assembly of the tetramer. Purified truncated subunits form aggregates that do not bind to blood group A+H substance (Etzler *et al.*, 1981). The intact subunits expressed in *E. coli* form oligomers with an anomalous molecular mass (Chao & Etzler, 1994). Crystals grown from recombinant DBL containing only intact subunits contain a mixture of intact and truncated subunits (Dao-Thi *et al.*, 1998).

The side chains of residues Ser 246 and Arg 250 of both helices protrude into a large, water accessible cavity situated in the center of the tetramer. The side chain of Arg 250 is completely disordered and is not visible in the electron density. On the opposite side of the α-helices, the side chains of residues Leu 244 and Leu 248 form the bottom of a large, hydrophobic cavity that contains density for two neighboring adenine residues (see next paragraph). For both tetramers in the asymmetric unit, the two intact subunits belong to a single canonical dimer (dimer AB in Fig. 4.1b) and the two truncated subunits belong to a second canonical dimer (dimer CD in Fig. 4.1b). A possible explanation for this fact is that an alternative arrangement (i.e., two canonical dimers each consisting of an intact and a truncated subunit) would bring the two Arg 250 side chains in the vicinity of each other, thereby creating an unfavorable interaction.

For the blood group A trisaccharide complex, one canonical dimer is present in the asymmetric unit and the complete tetramer is generated by the C2 symmetry operations. Similarly, for the Forssman disaccharide complex the complete tetramer is generated by applying the $I4_122$ symmetry operations to the single monomer in the asymetric unit. This means that in both cases the electron density in the center of the tetramer is averaged because of statistical disorder in the crystal. Therefore, two C-terminal stretches (Asp 236-Leu 253) with half occupancy are present in the former case, while one C-terminal stretch with half occupancy is present in the latter case.

4.3.2 Structure of the adenine binding site

Two identical adenine binding cavities with a hydrophobic character are found at opposite ends of the tetramer (Fig. 4.1b). The top of this cavity is formed by the $\beta\beta$-interface (Ile 189), while the bottom is formed by

Table 4.2: Hydrogen bonds and average distances between hydrogen bond donors and acceptors between DBL and adenine for the eight equivalent adenine binding sites in the asymetric unit. Three of the four hydrogen bond partners belong to the same subunit; the asterisk in front of Ser 178 indicates it belongs to the facing subunit (Fig. 4.2b).

Adenine atoms	DBL	Distance (Å)
N1	*Ser 178 Oγ	3.3
N3	Ser 178 Oγ	3.3
N6	Leu 165 O	3.5
N6	Thr 167 Oγ	2.8
N7	Thr 167 Oγ	2.9

Table 4.3: List of DBL residues with atoms within 4.6 Å of an adenine molecule in each of the four equivalent adenine binding sites. Most residues belong to one subunit; an asterisk in front of a residue indicates it belongs to the facing subunit. The number of atoms within 4.6 Å is shown between brackets.

Adenine atom	DBL residue (Number of atoms within 4.6 Å)
N1	Leu 165 (5), Val 176 (4), Ser 178 (3),
	Ala 177 (1)
C2	Leu 165 (2), Val 176 (1), Ser 178 (2),
	Ile 189 (1), *Leu 165 (1), *Ser 178 (1)
N3	Leu 165 (1), Val 176 (1), Ser 178 (1),
	Ile 189 (1), *Leu 165 (1), *Ser 178 (2)
C4	Val 176 (1), *Leu 165 (1), *Ser 178 (1)
C5	Leu 165 (1), Thr 167 (3), Val 176 (3)
C6	Leu 165 (3), Thr 167 (4), Val 176 (5)
N7	Thr 167 (3), Val 176 (3)
C8	Thr 167 (1)
N9	*Leu 165 (2), *Ser 178 (1), *Val 180 (1)

side chains that protrude from the sandwiched α-helices (Leu 244, Leu 248) (Fig. 4.2a). The side chains of the two C-terminal residues (Val 252, Leu 253) also add to the hydrophobic character of the cavity. The walls are formed by the protruding side chains from the back sheets of a pair of facing monomers (monomer pairs AC and BD in Fig. 4.1b). All the side chains that form the cavity are aliphatic (Leu, Val, Ile) or contain a hydroxyl group (Ser, Thr). The two cavities each have internal pseudo two-fold symmetry (broken by the sandwiched α-helices), and they are related to each other by a non-crystallographic two-fold axis. Each cavity contains electron density for two neighboring adenine molecules. The adenine molecules are bound by hydrogen bonds (Leu 165, Thr 167, Ser 178; see Table II) and hydrophobic interactions (Leu 165, Val 176, Val 180, Ile 189; see Table III) with side chains from β-strands 4, 5 and 6 of the back sheet (Fig. 4.2a and 4.2b).

In addition, two residues (Leu 244, Leu 248) belonging to the sandwiched α-helix are directly below the adenine rings. All nitrogen atoms of

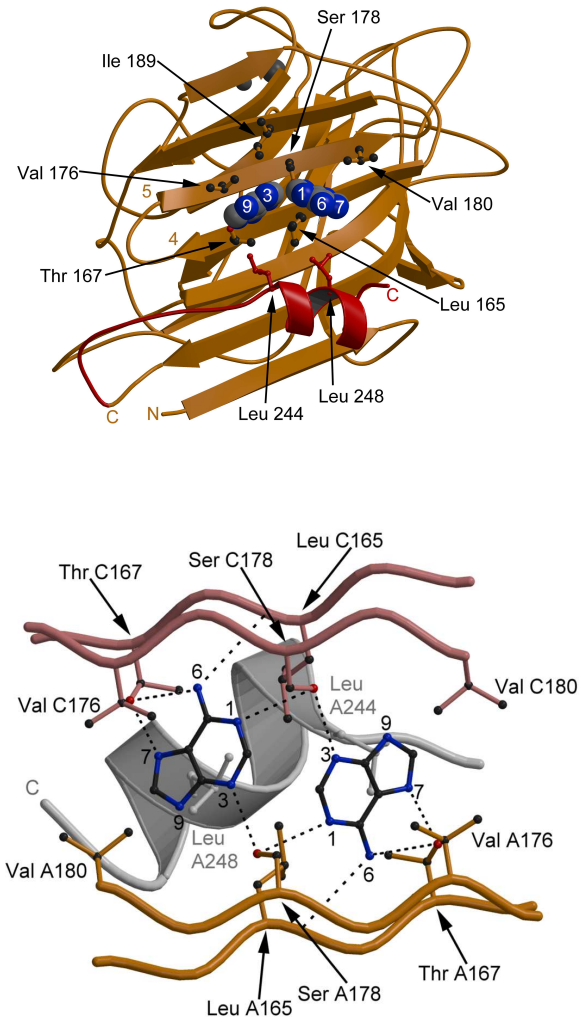


Figure 4.2: (Top) A view on a pair of neighboring adenine binding sites. The six β-strands that form the back sheet are numbered. The complete adenine binding site is formed by corresponding residues from two facing monomers (monomers pairs AC and BD in Fig. 4.1b). In this figure only the side chains of the residues involved in adenine binding from one monomer (monomer A in Fig. 4.1b) and the sandwiched α-helix (belonging to monomer A) are shown as ball-and-stick representations. (Bottom) A detailed view from the $\beta\beta$-interface of the DBL tetramer on the two neighboring adenine binding sites in dimer AC from Fig 4.1b. The upper two stran ds belong to monomer C, the lower two and the α-helix belong to monomer A. All residues involved in hydrogen bonds or hydrophobic interactions with an adenine molecule are shown as ball-and-stick models and labeled, except Ile A189 and Ile C189 which are not shown for clarity. The sandwiched α-helix that forms the bott om of the adenine binding site is shown in gray.

the adenine molecules are involved in hydrogen bonds, except N9. Weak density in the neighborhood of the N9 atoms indicates that it is probably involved in a water bridge with the main chain oxygen of Glu 136. The position of the adenine binding site is in accordance with the results of photoaffinity labeling experiments (Maliarik & Goldstein, 1988; Gegg & Etzler, 1994) and the putative location of the adenine binding site in PHA-L and SBA as previously suggested by us (Hamelryck *et al.*, 1996).

We observe four adenine molecules with half occupancy, in accordance with binding studies indicating that DBL binds two adenine molecules per tetramer (Gegg *et al.*, 1992). In each cavity, two facing Ser 178 residues are in the vicinity of the two neighboring adenine molecules. The position of each Ser 178 side chain oxygen is compatible with a role as hydrogen bond donor to either the N1 atom of one adenine residue, or the N3 atom of the neighboring adenine residue (Fig. 4.2b). The average distance in both cases is 3.3 Å. Hydrogen bond donor and acceptor sites in adenine are well defined (Jeffrey & Saenger, 1991). Nitrogens N1 and N3 only act as hydrogen bond acceptors and since the Ser residue can only serve as a donor in one hydrogen bond, it cannot be involved in these two hydrogen bonds at the same time. This explains why only two adenine molecules bind per tetramer, despite the presence of four (two in each cavity) potential adenine binding sites. Upon binding of one adenine molecule in a cavity, one Ser 178 hydroxyl group hydrogen bonds to N1, while the other Ser 178 residue hydrogen bonds to N3. In the same cavity, a second adenine molecule then necessarily binds with a much lower affinity because two hydrogen bond donors are unavailable. Hence, each cavity will only bind one adenine molecule, bringing the number of bound adenine molecules per tetramer to two. Due to statistical disorder in the crystal, electron density is present for four adenine molecules, instead of two.

The two neighboring adenine binding sites in a cavity are not completely equivalent. The two Leu side chains that protrude from the sandwiched α-helix (Leu 244, Leu 248), and form the bottom of the adenine binding sites, are each in the vicinity of one adenine molecule (Fig. 4.2b). However, the side chain of Leu 244 is considerably closer to the plane of the five-membered ring of the adenine molecule than the side chain of Leu 248 (4.5 Å and 5.5 Å, respectively).

4.3.3 DB58, a vegetative legume lectin from *D. biflorus*

Our results indicate that adenine binding by a legume lectin depends on the tetrameric structure shared by PHA-L, SBA and DBL. However, the stems and leaves of *Dolichos biflorus* also contain a 58 kD dimeric adenine binding legume lectin (DB58) (Etzler, 1994a). DB58 is a heterodimer of an intact (253 residues) and a truncated subunit (241 or 242 residues), and has 87 % sequence identity with DBL. DB58 proved to be very difficult to crystallize, but a 3.3 Å dataset of uncomplexed DB58 (Dao-Thi *et al.*, 1998) was collected at the Daresbury synchrotron facility allowing us to describe the quaternary structure of DB58. Three DB58 dimers are present in the asymmetric unit. Surprisingly, two of these dimers associate in the crystal to form a tetramer that resembles a DBL tetramer. However, the third DB58 dimer does not form a tetramer in the crystal and thus represents the true quaternary structure of the dimer in solution. DB58 forms a novel dimer that is similar to the top half of a DBL tetramer (monomer pairs AC and BD in Fig. 4.1b), i.e., it corresponds to two facing monomers that form

Figure 4.3: The overall structure of the DB58 dimer. The dimer corresponds to dimer AC in the DBL tetramer (Fig. 4.1b). Metal ions as in Fig. 4.1a. The positions of the conserved monosaccharide binding sites are indicated with asterisks.

the adenine binding site (Fig. 4.3). In DB58, the α-helix breaks the potential 2-fold symmetry of the dimer. Tetramer formation by DB58 in solution is probably hindered by the substitution of Pro 14 in DBL by Ser 14 in DB58. In the canonical dimers present in DBL, the Pro 14 ring is involved in a hydrophobic inter-subunit contact with Tyr 203. At present, three different tetramer types and three different dimer types have been described for the legume lectins (Fig. 4.4). The DB58 dimer brings the number of known legume lectin dimer types to four. The DBL tetramer can thus be considered as a dimer of DB58 dimers, that associate with each other via the formation of two canonical dimers in the tetramer. Similarly, the peanut agglutinin dimer can be considered as a dimer of two *Griffonia simplicifolia* lectin IV dimers that associate with each other via the formation of one canonical dimer. The quaternary structure of a legume lectin becomes important in binding multivalent ligands, which can lead to the formation of ordered, cross-linked lattices (Brewer, 1996). Our results indicate that the quaternary structure of certain legume lectins is also important for the formation of a hydrophobic binding site at a subunit interface.

All the residues involved in adenine binding in DBL are fully conserved in DB58, with the exception of the conservative substitution of Leu 244 by Ile. Formation of a DB58 dimer buries an area of 1800 Å^2, which is comparable to the buried area in a canonical dimer (1650 Å^2). The fact that both DBL and DB58 bind adenine and adenine derived plant hormones, despite their different quaternary structure, carbohydrate binding specificity (Etzler, 1994b) and locus of expression *in vivo* (Roberts & Etzler, 1984), illustrates the high potential importance of this binding site.

4.3.4 Blood group A trisaccharide binding

For the blood group A trisaccharide complex, strong electron density is only found for the GalNAc moiety of the trisaccharide, which indicates that the Fuc(α1–2)Gal moiety is not tightly bound by DBL. The visible GalNAc is bound in the conserved monosaccharide binding site, next to the equally

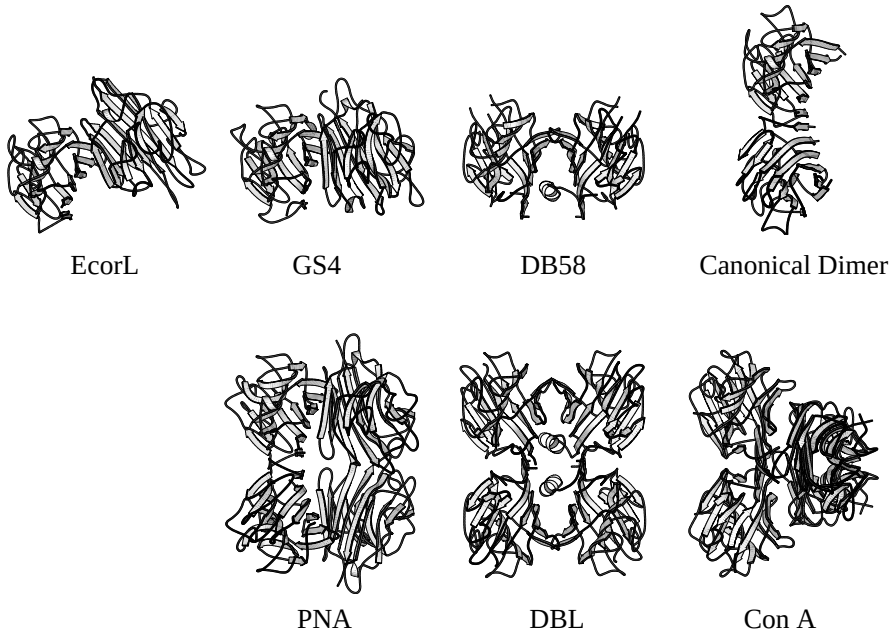


Figure 4.4: The different legume lectin oligomer types, including the novel DB58 dimer type (EcorL: *Erythrina corallodendron* lectin; GS4: *Griffonia simplicifolia* lectin IV; PNA: Peanut agglutinin; DB58: *Dolichos biflorus* stem and leaf lectin; DBL: *Dolichos biflorus* seed lectin; Con A: Concanavalin A). The canonical dimer type is represented by the lentil lectin dimer. Each tetramer type can be considered as a dimer of dimers. In the figure, the dimers and the corresponding tetramers are vertically aligned. No tetramer is known that contains an EcorL dimer. One of the subunits is in the same orientation for all oligomers (upper left monomer for the three tetramers). β-Strands are shown as arrows.

conserved metal binding site (Fig. 4.5a). Three conserved sugar binding residues (Asp 85, Gly 103 and Asn 129) occupy positions similar to the other legume lectin structures. The Asp 85 residue, which is involved in the *cis*-peptide bond with the preceding Ala residue, hydrogen bonds to the hydroxyl groups at positions 3 and 4 of the GalNAc residue. Asn 129, which interacts with the bound Ca^{2+} ions via its Oδ^1 atom, hydrogen bonds to the hydroxyl group at position 4 via its Nδ^2 atom. Gly 103 plays a crucial role in the GalNAc-DBL complex. The backbone amide group of Gly 103 is involved in a bifurcated hydrogen bond with O3 and oxygen of the N-acetyl group. Further hydrogen bonds are formed between Ser 215 Oγ and O6 and between Leu 214 N and O4. Leu 214 is responsible for the a-anomeric preference of DBL. Binding of β-GalNAc in the monosaccharide binding site of DBL would bring the β-anomeric oxygen close to the side chain of Leu 214.

Leu 214 and Ser 215 belong to the so-called specificity loop (Sharma & Surolia, 1997; Loris *et al.*, 1998). While the conserved Asn-Gly-Asp triad and a conserved aromatic residue (see below) confer affinity, the specificity loop is thought to confer specificity as well as affinity by excluding certain monosaccharides via sterical hindrance while interacting favorably with others (Sharma & Surolia, 1997; Loris et al., 1998). Hydrogen bonds between protein and sugar are listed in Table IV.

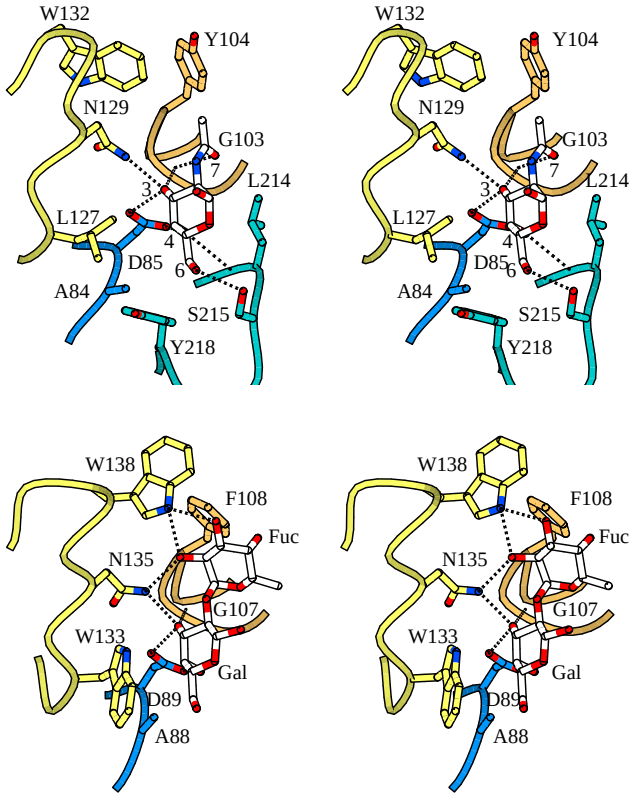


Figure 4.5: (Top) A view on the monosaccharide binding site of the DBL-blood group A trisaccharide complex. The visible GalNAc residue and the side chains of sugar binding residues are shown as ball-and-stick models. The sugar binding residues belong to four different stretches, each shown in a different color (metal loop: yellow; *cis*-peptide bond region: blue; specificity loop: green; Ω-loop: orange). Nitrogen atoms are shown in blue; oxygen atoms in red. Hydrogen bonds between sugar and protein residues are shown as dotted lines. The sugar hydroxyl groups that are involved in hydrogen bonds with the protein are labeled. (Bottom) GS4 in complex with the Le^b tetrasaccharide. For clarity, only the Fuc(α1–2)Gal moiety is shown. The view on the sugar binding site of GS4 correspond s to the view on the sugar binding site of DBL in Fig. 4.5a. Sugar binding residues are shown as ball-and-stick models. Atom color coding, hydrogen bonds and loop colors as in Fig. 4.5a.

Table 4.4: Hydrogen bonds and average distances between hydrogen bond donors and acceptors between the non-reducing GalNAc in the monosaccharide binding site and DBL for the blood group A trisaccharide complex and the Forssman disaccharide complex.

Hydroxyl group	Protein atom	Distance (Å)
O3	Asp 85 Oδ1	2.9
	Gly 103 N	2.9
	Asn 129 Nδ2	3.1
O4	Asp 85 Oδ2	2.6
	Leu 214 N	3.1
O6	Ser 215 Oγ	2.7
O7	Gly 103 N	3.0

In addition to hydrogen bonds, hydrophobic interactions also play an important role: the side chains of Tyr 104 and Trp 132 form a hydrophobic pocket which is in the vicinity of the methyl group of GalNAc (atoms of both residues within 4.6 Å from the carbon atom are: Tyr 104 Cδ^2: 4.3 Å; Tyr 104 Cε^2: 3.8 Å; Trp 132 Cη^2: 4.1 Å). Binding of GalNAc buries 23 Å^2 of the non-polar area of both residues. Tyr 104 belongs to a large W-loop, together with the conserved Gly 103 residue, while Trp 132 is part of the metal binding loop. Tyr 218, which belongs to the specificity loop, makes a favorable hydrophobic interaction with the C6 atom. The interaction buries 11 Å^2 of the non-polar area of Tyr 218. Furthermore, the Leu 127 residue packs against the hydrophobic patch of the B face of the GalNAc moiety, formed by C3, C4, C5 and C6. This buries 16 Å^2 of the non-polar area of Leu 127. In most legume lectins, an aromatic residue is present in the position of Leu 127, and the significance of this substitution for the specificity of DBL is discussed in the next paragraph. All DBL residues with atoms within 4.6 Å of the bound sugar are listed in Table V.

Imberty *et al.* (1994) presented a model of the DBL-Blood Group A trisaccharide complex that is in agreement with NMR studies (Casset et al., 1996). The structure of the conserved sugar binding site and the position of the GalNAc residue in this site are in perfect agreement with the here presented crystal structure. In the model, the Fuc(α1–2)Gal disaccharide wraps around Leu 127 and the Fuc moiety hydrogen bonds to the carbonyl group of Ser 128. In the crystal structure, some weak density is indeed seen for the Fuc moiety in the neighborhood of Ser 128, indicating that it probably transiently binds to the lectin in the proposed Fuc binding subsite.

4.3.5 GalNAc specificity

4.3.5.1 The role of Leu 127

DBL is a strict GalNAc specific lectin (K_a=4.2 10^3 M^{-1} for Me-α-D-GalNAc, Etzler *et al.*, 1981), in the sense that precipitation of blood group A+H substance by DBL is inhibited by GalNac but not by Gal (Etzler & Kabat, 1970). To determine the molecular basis underlying the specificity of DBL, the sugar binding site of DBL was compared with the binding sites of four Gal binding legume lectins (EcorL (Shaanan et al., 1991), SBA (Dessen *et al.*, 1995), PNA (Banerjee *et al.*, 1994) and WBAI (Prabu *et al.*, 1998)). In

Table 4.5: DBL residues with atoms within 4.6 Å of a carbohydrate atom. An asterisk in front of the residue name indicates it belongs to a symmetry mate. The number of atoms within 4.6 Å is shown between brackets.

Non-reducing GalNAc (blood group A trisaccharide and Forssman disaccharide)	
Sugar ring (C1, C2, C3, C4, C5, O5)	Asp 85 (6), Gly 103 (2), Asn 129 (2), Leu 214 (10), Leu 217 (2)
N-acetyl group (N2, C7, O7)	Asn 101 (2), Gly 102 (4), Gly 103 (5), Asn 129 (2), Leu 214 (3)
N-acetyl group (C8)	Gly 103 (2), Tyr 104 (2), Asn 129 (1), Trp 132 (1)
O3	Asp 85 (3), Gly 102 (2), Gly 103 (2), Leu 127 (1), Asn 129 (3)
O4	Asp 85 (3), Gly 102 (1), Leu 127 (1), Gly 213 (3), Leu 214 (5)
C6 and O6	Gly 213 (2), Leu 214 (5), Ser 215 (8), Tyr 218 (11)
Reducing GalNAc (Forssman disaccharide)	
N-acetyl group (C8)	Leu 127 (2), Tyr 218 (3), *Asp 109 (6), *Asn 114* (1)

EcorL, SBA, PNA and WBAI there is an aromatic residue (Phe 128 in SBA, Phe 131 in EcorL, Tyr 125 in PNA, Phe 126 in WBAI) that stacks against the C3, C4, C5 and C6 patch of Gal. In DBL, this residue is replaced by the aliphatic residue Leu 127. Strikingly, in all known crystal structures of lectins complexed with Gal, the apolar C3-C4-C5-C6 patch of the B face of the sugar packs against an aromatic residue (Rini, 1995; Weis & Drickamer, 1996), clearly demonstrating the importance of this interaction. Interactions of Man with aromatic residues show much more variability and seem less critical. This leads to the conclusion that the favorable interactions between the N-acetylgroup and the Gly 103/Tyr 104/Trp 134 triad compensate for the substitution of an aromatic residue by an aliphatic residue (Leu 127). Since only GalNAc is able to compensate for the replacement of an aromatic residue with Leu, Gal binds with a much lower affinity.

To assess the role of the Leu 127 residue, the mutant Leu127Phe was constructed and binding studies were performed. The Leu127Phe mutant binds to hog blood A+H substance conjugated to Sepharose beads better than the native seed lectin (Fig. 4.6). Fifty percent inhibition of binding for the native lectin occurred at 120.2 mM Gal and 1.1 mM GalNAc. For the Leu127Phe mutant, fifty percent inhibition of binding occurred at only 28.4 mM Gal and 0.27 mM GalNAc. These results confirm that the Leu 127 residue is one of the determining factors of the low affinity of DBL for Gal.

The results from two other studies on the effect of mutations on sugar binding by two Gal binding legume lectins are in full accord with this view. In EcorL, two mutants in which the conserved aromatic residue is mutated to an aliphatic residue (Phe131Ala and Phe131Val) cannot induce hemagglutination (Adar & Sharon, 1996). In the *Robinia pseudoacacia* lectin, the Phe130Ala and the Phe130Leu mutations impair hemagglutination as

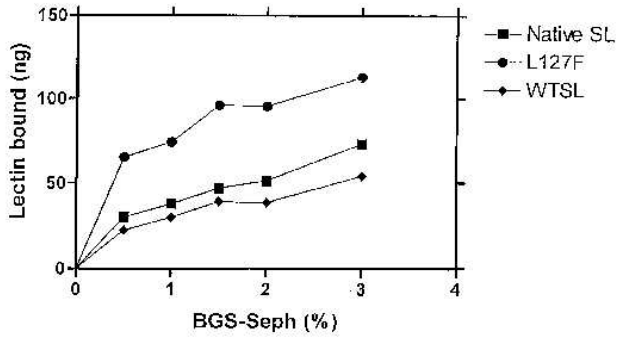


Figure 4.6: Binding of iodinated lectins (wild type DBL, WTSL; native recombinant DBL, Native SL; and the Leu127Phe mutant, L127F) to hog blood group A+H Sepharose.

well (Nishiguchi *et al.*, 1997). This again demonstrates that an aliphatic residue (Val, Leu) cannot compensate for the loss of a sugar-stacking aromatic residue.

DBL is not the only legume lectin in which the conserved aromatic residue that stacks against the sugar ring is replaced by an aliphatic residue. The structure of (uncomplexed) PHA-L (Hamelryck *et al.*, 1996) from the common bean reveals that this lectin has a Leu residue (Leu 126) at this position. This may explain why precipitation by PHA-L is not inhibited by monosaccharides (Hammarström *et al.*, 1982).

4.3.5.2 The role of the specificity loop

Recently, the crystal structure of EcorL in complex with GalNAc was published (Elgavish & Shaanan, 1998). Unlike DBL, EcorL binds Gal and GalNAc with equal affinity (K_a=2.4 M^{-1} for Gal; K_a=2.3 M^{-1} for GalNAc) (Surolia *et al.*, 1996). To study the molecular basis of the difference in specificity between EcorL and DBL, the monosaccharide binding sites of DBL and EcorL were superimposed using the Cα atoms of the four conserved sugar binding residues (Asp 85, Leu 127, Gly 103 and Asn 129 for DBL; Asp 89, Phe 131, Gly 107 and Asn 133 for EcorL) (Fig. 4.7). The four residues occupy virtually identical positions in both lectins (rmsd=0.1 Å). Surprisingly, the three residues that interact with the N-acetyl group of GalNAc in DBL (Gly 103, Tyr 104 and Trp 132) are present in EcorL in the same position (Gly 107, Tyr 108, Trp 135). The absence of a higher affinity for GalNAc in EcorL is thus not due to the simple absence of the Gly-Tyr-Trp triad.

The position of the GalNAc residue in the monosaccharide binding sites of DBL and EcorL is similar, but not identical. In EcorL, the sugar ring is about 0.8 Å closer to the metal binding loop than in DBL. Again, the monosaccharide specificity loop plays an important role in positioning the monosaccharide in the monosaccharide binding site. In EcorL, Ser 215 and Tyr 218 are replaced by Gln 219 and Ala 222. These two substitutions position the sugar ring closer to the metal binding site in EcorL than in DBL, and push the sugar ring against the aromatic ring of Phe 131. In DBL, the side chains of Ser 215 and Tyr 218 position the sugar closer to the specificity loop. A second difference is the presence of Tyr 106 in EcorL,

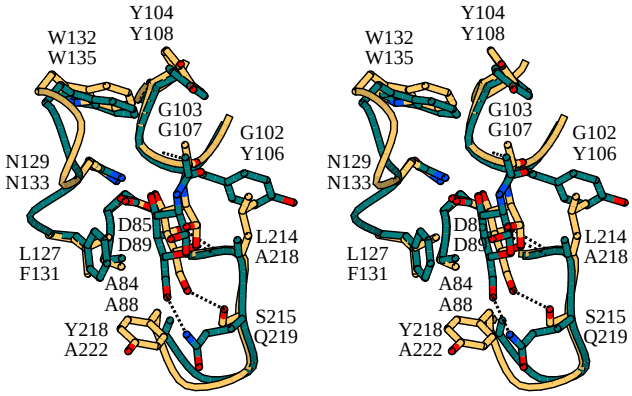


Figure 4.7: Stereo figure of the superimposed monosaccharide binding sites of DBL (yellow) and EcorL (green). All residues involved in sugar binding are shown as ball-and-stick models. The sugar binding residues are labeled for DBL (upper) and EcorL (lower). Atom color coding and hydrogen bonds as in Fig. 4.5.

which corresponds to Gly 102 in DBL. The Cβ atom of this residue might push the N-acetyl group away from the conserved Gly residue in EcorL. As a consequence of the above discussed structural features, the distance between the amide group of Gly 107 and the oxygen atom of the N-acetyl group of GalNAc in EcorL (4.0 Å) is 1.1 Å longer than the corresponding distance in DBL (2.9 Å). Thus, EcorL does not bind GalNAc with a higher affinity than Gal because the sugar ring is not optimally positioned in the monosaccharide binding site.

4.3.6 Forssman disaccharide binding

The quality of the electron density of the Forssman disaccharide was high for both residues (Fig. 4.8), in contrast to the absence of electron density for two of the three residues of the blood group A trisaccharide. No density is seen for the O6 hydroxyl group of the reducing GalNAc residue: this hydroxyl group is not in contact with the protein and seems to be completely flexible. The group attached to O1 of the reducing GalNAc extends into the solvent and is not visible in the electron density.

The non-reducing GalNAc residue of the GalNAc(α1–3)GalNAc disaccharide is bound in the monosaccharide binding site in exactly the same way as the GalNAc residue of the blood group A trisaccharide (Fig. 4.9, Table IV, Table V). The reducing GalNAc residue extends from the monosaccharide binding site and only interacts with the protein via its N-acetyl group (Table V). The methyl group of the N-acetyl moiety of the reducing GalNAc residue points to a hydrophobic cavity formed by the side chains of residues Leu 127 and Tyr 218. Both residues are also involved in binding the non-reducing GalNAc in the monosaccharide binding site. It should also be noted that the methyl group of the reducing GalNAc is located directly above the hydrophobic patch (distances are 4.0 Å from C5; 4.5 Å from C6) of the non-reducing GalNAc. Binding of the Forssman disaccharide therefore creates a small, solvent shielded hydrophobic cavity consisting of the side chains of Leu 127 and Tyr 218, the methyl group of the reducing GalNAc and the hydrophobic patch of the non-reducing

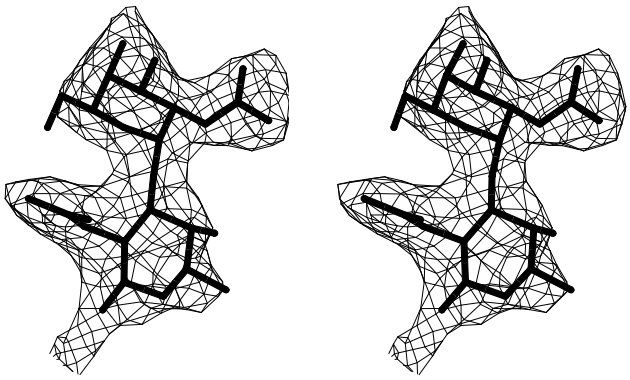


Figure 4.8: Stereo figure of the F_o-F_c, Φ_c density around the Forssman disaccharide, calculated after simulated annealing refinement of the DBL structure in the absence of the sugar. The density was visualized at a level of 1.0 σ.

GalNAc. The presence of a Leu residue at position 127 in DBL, instead of an aromatic residue as in most legume lectins, plays an important role in Forssman disaccharide binding: the larger Phe, Tyr or Trp side chains would clash with the N-acetyl group.

Binding of GalNAc in the monosaccharide binding site buries 27 Å^2 of the non-polar area of Leu 127 and Tyr 218, and the reducing GalNAc residue buries another 18 Å^2. No additional hydrogen bonds are made with the protein by the reducing GalNAc residue, and the shielding of the hydrophobic cavity from the solvent appears to be the crucial specificity-determining interaction between the Forssman disaccharide and DBL. The crucial role of the N-acetyl group in enhancing the affinity is further confirmed by the fact that Me-α-D-GalNAc and GalNAc(α1–3)Gal are equally active as inhibitors of precipitation of blood group A substance (Etzler & Kabat, 1970).

The conformation of the Forssman disaccharide in the crystal structure (Ψ=88°; Θ=70°) broadly resembles the conformation in the modeled complex (Y=77°; Θ=80°) (Imberty *et al.*, 1994). This conformation corresponds to the lowest energy conformation in molecular mechanics simulations and is in accordance with NMR measurements of the disaccharide in solution (Grönberg *et al.*, 1994; Casset *et al.*, 1996) and bound to DBL (Casset *et al.*, 1997; Rinnbauer *et al.*, 1998). These observations support the view that lectins generally bind sugars in a conformation that is close to a minimum energy conformation in solution.

4.3.7 A multi-purpose secondary binding site in the legume lectin family

The N-acetylgroup of the GalNAc residue bound in the monosaccharide binding site of DBL interacts with a hydrophobic cavity formed by Tyr 104 and Trp 132. Comparison of this N-acetyl binding region in DBL with the corresponding region in GS4 (Delbaere *et al.*, 1990) reveals that they both possess a similar hydrophobic pocket, formed by Tyr 104 and Trp 132 in DBL and Phe 108 and Trp 138 in GS4 (Fig. 4.5b). GS4 belongs to the complex specificity group and does not bind monosaccharides. How-

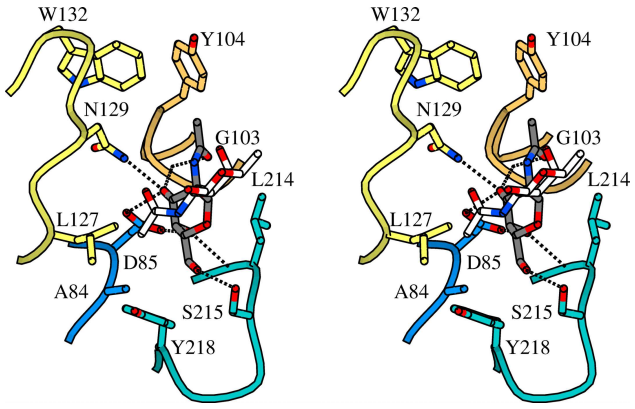


Figure 4.9: Stereo figure of the Forssman disaccharide in complex with DBL. The non-reducing GalNAc residue bound in the conserved monosaccharide binding site is shown in gray; the reducing GalNAc residue is shown in white. Sugar binding residues are shown as ball-and-stick models. The view shown in this figure is the same as in Fig. 4.5. Atom color coding, hydrogen bonds and loop colors as in Fig. 4.5.

ever, the complex of GS4 with the Le^b oligosaccharide (Fuc(α1–2)Gal(β1–3)[Fuc(α1–4)]GlcNAc) shows that the Gal residue is bound in a truncated monosaccharide binding site, in a way similar to Gal in Gal specific legume lectins. In the GS4-Le^b complex, the Fuc residue (α1–2) linked to the Gal residue in the monosaccharide binding site is bound in the region that corresponds to the N-acetyl binding pocket in DBL. The Fuc residue is involved in a stacking interaction with Phe 108 and its O2 and O3 hydrogen bond to Trp 138 Nε^1. The latter residue corresponds to Trp 132 in DBL, but adopts a different conformation in GS4.

In the modeled structure of lentil lectin complexed with 3-O-nitrophenyl-mannose, the same subsite harbors the nitrophenylgroup (Loris *et al.*, 1994) which interacts with the side chains of Tyr 100β and Trp 128β, corresponding to Tyr 104 and Trp 132 in DBL. In the modeled structure of EcorL with Fuc(α1–2)Gal(β1–4)GlcNAcβ (Moreno et al., 1997), again the same subsite interacts with the fucose moiety, involving the side chains of Phe 108 and Trp 135. In both structures, the backbone amide groups of the Gly residues equivalent to Gly 103 in DBL are involved in a bifurcated hydrogen bond (with O3 of the mannose moiety in lentil lectin, and O2 of fucose in EcorL). Mutagenesis experiments and molecular modeling have indicated that the same subsite is also involved in the high affinity binding of methyl-α-N-dansylgalactosaminide by EcorL (Arango *et al.*, 1993; Adar et al., 1998). The site formed by the Gly-Tyr-Trp triad thus seems to be a conserved multi-purpose binding site, involved in binding both hydrophobic groups and monosaccharide residues.

4.4 Conclusions

The structures of DBL in complex with adenine and of uncomplexed DB58 show that certain legume lectins possess a hydrophobic binding site which depends on their quaternary structure. The biological relevance of this adenine/cytokinin binding site is unclear, but the fact that it is conserved

in (at least) one dimeric (DB58) and three tetrameric (SBA, PHA-L and DBL) legume lectins from different plant species indicates that it might play an important biological role. Both DBL and DB58 possess an unusual architecture, featuring an α-helix sandwiched between two monomers. In both oligomers, this α-helix is directly involved in the formation of the hydrophobic binding site. DB58 adopts a novel quaternary structure, related to the quaternary structure of the DBL heterotetramer, and brings the number of know legume lectin dimer types to four.

The DBL sugar complexes illustrate that specific binding of an N-acetylated sugar is achieved through a give-and-take mechanism: a lack of aromatic stacking against the sugar ring is compensated by a favorable interaction of the N-acetyl group with the lectin. The latter interaction not only depends on the presence of a specific subsite, but also on the correct positioning of the sugar ring by the specificity loop. High affinity binding of the Forssman disaccharide is solely mediated via the formation of a small, solvent shielded hydrophobic cavity and does not involve any additional hydrogen bonds.

4.5 References

Adar, R., Moreno, E., Streicher, H., Karlsson, K.A., Ångström, J. & Sharon, N. (1998) Protein Science, 7, 52-63.

Adar, R. & Sharon, N. (1996) Eur. J. Biochem., 239, 668-674.

Arango, R., Rodriguez-Arango, E., Adar, R., Belenky, D., Loontiens, F.G., Rozenblatt, S. & Sharon, N. (1993) FEBS lett., 330, 133-136.

Baker, D.A., Sugii, S., Kabat, E.A., Ratcliffe, R.M., Hermentin, P. & Lemieux, R.U. (1983) Biochemistry, 22, 2741-2750.

Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Surolia, A. & Vijayan, M. (1994) Proc. Natl. Acad. Sci. USA, 91, 227-231

Brewer, C.F. (1996) Chemtracts Biochem. Mol. Biol., 6, 165-179

Brünger, A. (1992). X-PLOR version 3.1 : A system for crystallography and NMR. Yale University, New Haven, CT.

Carter, W.G. and Etzler, M.E. (1975) J. Biol. Chem., 250, 2756-2762.

Casset, F., Peters, T., Etzler, M., Korchagina, E., Nifant'ev, N., Pérez, S. & Imberty, A. (1996) Eur. J. Biochem., 239, 710-719.

Casset, F., Imberty, A., Pérez, S., Etzler, M.E., Paulsen, H. & Peters, T. (1997) Eur. J. Biochem., 244, 242-250.

Chao, Q., Casalongue, C., Quinn, J.M. & Etzler, M.E. (1994) Arch. Biochem. Biophys., 313, 346-350.

Collaborative computational project, n.4., SERC Daresbury laboratory. (1994) Acta Cryst., D50, 760-763.

Connolly, M.L. (1993) J. Mol. Graphics, 11, 139-141.

Cowtan, K.D. & Main, P. (1993) Acta Cryst., D49, 148-157.

Dao-Thi, M., Hamelryck, T.W., Bouckaert, J., Körber, F., Burkow, V., Poortmans, F., Etzler, M., Strecker, G., Wyns, L., & Loris, R. (1998) Acta Cryst., D54, 1446-1449

Delbaere, L., Vandonselaar, M., Prasad, L., Quail, W., Pearlstone, J., Carpenter, M., Smillie, L., Nikrad P., Spohr, U. & Lemieux, R. (1990) Can. J. Chem., 68, 1116-1121

Dessen, A., Gupta, D., Sabesan, S., Brewer, F. & Sacchettini, J.C. (1995) Biochemistry, 34, 4933-4942.

Elgavish, S. & Shaanan, B. (1998) J. Mol. Biol., 277, 917-932.

Esnouf, R.M. (1997) J. Mol. Graphics, 15, 132-134.

Etzler, M.E. (1994a) Biochemistry, 33, 9778-9783.

Etzler, M.E. (1994b) Glycoconjugate J., 11, 395-399.

Etzler, M.E. (1996). In Lectins, Biology-Biochemistry-Clinical Biochemistry. (Van Driessche, E., Rougé, P., Beeckmans, S. & BÆg-Hansen, T.C., eds), pp. 3-9, Textop. Hellerup, Denmark

Etzler, M.E., Gupta, S. & Borrebaeck, C. (1981) J. Biol. Chem., 256, 2367-2370.

Etzler, M.E. & Kabat, E.A. (1970) Biochemistry, 9, 869-877.

Gegg, C.V. & Etzler, M.E. (1994) J. Biol. Chem., 269, 5687-5692.

Gegg, C.V., Roberts, D.D., Segel, I.H. & Etzler, M.E. (1992) Biochemistry, 31, 6938-6942.

Grönberg, G., Nilsson, U., Bock, K. & Magnusson, G. (1994) Carbohydrate Res., 257, 35-54.

Hakomori, S. (1984) Ann. Rev. Immunol., 2, 103-126.

Hamelryck, T.W., Dao-Thi, M., Poortmans, F., Chrispeels, M.J., Wyns, L. & Loris, R. (1996) J. Biol. Chem., 271, 20479-20485.

Hammarström, S., Hammarström, M.L., Sundblad, G., Arnarp, J. & Lönngren, J. (1982) Proc. Natl. Acad. Sci. USA, 79, 1611-1615.

Hammarström, S., Murphy, L.A., Goldstein, I.J. and Etzler, M.E. (1977) Biochemistry, 16, 2750-2755.

Harada, J.J., Spadoro-Tank, J., Maxwell, J.C., Schnell, D.J. & Etzler, M.E. (1990) J. Biol. Chem., 265, 4997-5002.

Imberty, A., Casset, F., Gegg, C.V., Etzler, M.E. & Pérez, S. (1994) Glycoconjugate J., 11, 400-413.

Jeffrey, G.A. & Saenger, W. (1991) Hydrogen bonding in biological structures. Springer-Verlag, Berlin-Heidelberg.

Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991) Acta Cryst, A47, 110-119.

Kraulis, P.J. (1991) J. Appl. Cryst., 24, 946-950.

Landersjö, C., Weintraub, A., Ansaruzzaman, M., Albert, M.J. & Widmalm, G. (1998) Eur. J. Biochem., 251, 986-990.

Landt, O., Grunert, H.P. & Hahn, U. (1990) Gene, 96, 125-128.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993) J. Appl. Cryst., 26, 283-291.

Loris, R., Casset, F., Bouckaert, J., Pletinckx, J., Dao-Thi, M., Poortmans, F., Imberty, A., Pérez, S. & Wyns, L. (1994) Glycoconjugate J., 11, 507-517.

Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998) Biochem. Biophys. Acta, 1383, 9-36.

Maliarik, M.J., Roberts, D.D. & Goldstein, I.J. (1987) Arch. Biochem. Biophys., 255, 194-200.

Maliarik, M.J. & Goldstein, I.J. (1988) J. Biol. Chem., 263, 11274-11279.

Maliarik, M., Plessas, N.R., Goldstein, I.J., Musci, G. & Berliner, L.J. (1989) Biochemistry, 28, 912-917.

Mandal, D.K., Nieves, E., Bhattacharyya, L., Orr, G.A., Roboz, J., Yu, Q., Brewer, C.F. (1994) Eur. J. Biochem., 221, 547-553

Mathieu, M., Modis, Y., Zeelen, J.P., Engel, C.K., Abagyan, R.A., Ahlberg, A., Rasmussen, B., Lamzin, V.S., Kunau, W.H. & Wieringa, R.K. (1997) J. Mol. Biol., 273, 714-728.

Merrit, E.A. & Murphy, M.E.P. (1994) Acta Cryst., D50, 869-873.

Moreno, E., Teneberg, S., Adar, R., Sharon, N., Karlsson, K. & Angström, J. (1997) Biochemistry, 36, 4429-4437.

Navaza, J. (1994) Acta Cryst, A50, 157-163.

Nishiguchi, M., Yoshida, K., Sumizono, T. & Tazaki, K. (1997) FEBS lett., 403, 294-298.

Otwinowski, Z. & Minor, W. (1997) Methods Enzymol., 276, 307-326.

Prabu, M.M., Sankaranarayanan, R., Puri, K.D., Sharma, V., Surolia, A.,

Vijayan, M. & Suguna, K. (1998) J. Mol. Biol., 276, 787-796

Rini, J.M. (1995) Annu. Rev. Biophys., 24, 551-577.

Rinnbauer, M., Mikros, E. & Peters, T. (1998) J. Carbohydrate Chemistry, 17, 217-230.

Roberts, D.M., Walker, J. & Etzler, M.E. (1982) Arch. Biochem. Biophys., 218, 213-219.

Roberts, D.M. & Etzler, M.E. (1984) Plant Physiol., 76, 879-884.

Roberts, D.D. & Goldstein, I.J. (1983) J. Biol. Chem., 258, 13820-13824.

Schnell, D.J. & Etzler, M.E. (1987) J. Biol. Chem., 262, 7220-7225.

Schnell, D.J. & Etzler, M.E. (1988) J. Biol. Chem, 263, 14648-14653.

Shaanan, B., Lis, H. and Sharon, N. (1991) Science, 254, 862-866

Sharma, V. & Surolia, A. (1997) J. Mol. Biol., 267, 433-445.

Surolia, A., Sharon, N. & Schwartz, F.P. (1996) J. Biol. Chem., 271, 17697-17703.

Weis, W.I. & Drickamer, K. (1996) Annu. Rev. Biochem. 65, 441-473.

Williams, M.A., Goodfellow, J.M. & Thornton, J.M. (1994) Protein Science, 3, 1224-1235.

Young, M.N., Watson, D.C., Yaguchi, M., Adar, R., Arango, R., Rodriguez-Arango, E., Sharon, N., Blay, P.K.S. & Thibault, P. (1995) J. Biol. Chem., 270, 2563-2570.

Chapter 5

Summary

Phytohemagglutinin-L (PHA-L) is a legume lectin found in the seeds of the common bean. It is a tetrameric lectin that specifically binds to leukocytes and belongs to the complex specificity group (best binder is Gal(β1–4)GlcNAc(β1–2)[Gal(β1–4)GlcNAc(β1–6)]Man, minimal binder is GlcNAc(β1–2)Man). PHA-L is very toxic for human consumption because it binds to the gut and disrupts its normal functioning. Every year, many people suffer from PHA-L poisoning. The true cause of the illness is practically always erroneously attributed to bacterial food poisoning.

The structure of PHA-L was solved at a resolution of 2.8 ÅLike the concanavalin A tetramer, the PHA-L tetramer consists of two canonical dimers that pack against each other. In concanavalin A, the two dimers pack perpendicular against each other, while in PHA-L the two dimers associate in a parallel fashion, which creates a large cavity in the molecule. Some legume lectins, incuding PHA-L, also bind adenine and adenine derived plant hormones, i.e., cytokinins. The exact function of this binding site *in vivo* is unknown, but it might point to an involvement in plant hormone signalling or storage. Based on photo affinity labeling of the adenine binding site in two closely to PHA-L related lectins (PHA-E and lima bean lectin), the putative location of the two adenine binding sites was determined to lie in the central channel. A comparison of the sugar binding site of PHA-L with those of other legume lectins led to the conclusion that PHA-L does not bind monosaccharides due to the absence of aromatic stacking against the sugar ring.

Arcelin is a truncated and hence inactive legume lectin from the common bean that is closely related to PHA-L. Arcelin is only found in a few wild bean accessions and protects against predation by the Mexican bean weevil (*Zabrotes subfasciatus*). The mechanism behind the toxicity of arcelin is unknown.

The structure of arcelin-5 was solved at a resolution of 2.7 ÅAs expected, the architecture of arcelin-5 resembles that of a typical legume lectin, with two important differences. First, arcelin-5 crystallized as a monomer, while all known legume lectins are oligomers. Second, arcelin-5 misses the conserved metal binding site that stabilises an equally conserved *cis*-peptide bond between an Ala and an Asp residue. This *cis*-peptide bond is crucial for the sugar binding activity of a legume lectin. Surprisingly, arcelin-5 has an Ala-Tyr *cis*-peptide bond at this position, despite the absence of stabilising metal ions. All known arcelin sequences show an Ala-Tyr or Ala-Phe sequence at this position. Since non X-Pro *cis*-peptide bonds are rare and often play a functional or catalytical role,

an Ala-Tyr/Phe *cis*-peptide bond might be crucial for the toxic activity of the arcelins.

The *Dolichos biflorus* seed lectin (DBL) has an interesting specificity among the legume lectins because it binds GalNAc readily, while it binds Gal very weakly. Most legume lectins bind Gal and GalNAc with a comparable affinity. In addition, like PHA-L, DBL is an adenine/cytokinin binding tetramer.

DBL was crystallized with the blood group A trisaccharide GalNAc(α1–3)[Fuc(α1–2)]Gal (2.8 Å), the Forssman disaccharide GalNAc(α1–3)GalNAc (2.6 Å) and adenine (2.65 Å). For the blood group trisaccharide only the Gal moiety bound in the conserved monosaccharide binding site was visible in the density. Comparison with Gal binding legume lectins shows that DBL binds Gal with low affinity because it lacks aromatic stacking against the Gal ring. GalNac binds with higher affinity because its N-acetyl group compensates for the lack of aromatic stacking by interacting with Gly 103, Tyr 104 and Trp 132.

The non reducing GalNAc of the Forssman disaccharide is bound in the conserved monosaccharide binding site in exactly the same way as the GalNAc residue of the blood group A trisaccharide. The reducing GalNAc does not make any additional hydrogen bonds with the lectins, but creates a hydrophobic cluster with the side chains of Tyr 218 and Ley 127, C5 and C6 of the non reducing GalNAc and the methyl group of the reducing GalNAc. High affinity binding of the Forssman disaccharide is thus solely mediated by hydrophobic interactions.

The quaternary structure of DBL is identical to PHA-L, but in the case of DBL the center of the tetramer contains two C-terminal α-helices sandwiched between the β-sheets two facing canonical dimers. The α-helices form the bottom of four potential adenine binding sites, two of which can be occupied at the same time, near two subunit interfaces.

The stems and leaves of *Dolichos biflorus* also contain a dimeric legume lectin (DB58) that binds adenine. This is surprising since adenine binding by PHA-L and DBL depends on their quaternary structure. The structure of DB58 at 3.3 Å resolution shows that it adopts a novel quaternary structure in which the adenine binding site is conserved. DB58 is thus a representative of a fourth legume lectin dimer type.

Chapter 6

Samenvatting

Phytohemagglutinine-L (PHA-L) is een *Leguminosae* lectine dat voorkomt in de zaden van de gewone boon (*Phaseolus vulgaris*). Het is een tetrameer dat specifiek bindt op leukocyten en behoort tot de complexe specificiteitsgroep van *Leguminosae* lectines (beste binder is Gal(β1–4)GlcNAc(β1–2)-[Gal(β1–4)GlcNAc(β1–6)]Man, minimale binder is GlcNAc(β1–2)Man). PHA-L is sterk toxisch voor zoogdieren: het lectine bindt op het darmoppervlak en interfereert met de normale voedselopname. Jaarlijks belanden heel wat mensen in het ziekenhuis na het consumeren van onvoldoend gekookte bonen. Meestal wordt de PHA-L vergiftiging dan foutief toegeschreven aan bacteriële voedselvergiftiging.

De structuur van (ongecomplexeerd) PHA-L werd opgelost bij een resolutie van 2.8 ÅḢet tetrameer bestaat, net als het tetrameer van concanavaline A (Con A), uit twee kanonieke *Leguminosae* lectine dimeren. In tegenstelling tot Con A associëren de twee dimeren echter met elkaar op een quasi parallelle wijze, hetgeen een grote holte in het centrum van het tetrameer creëert. Sommige *Leguminosae* lectines, waaronder PHA-L, binden adenine en plantenhormonen afgeleid van adenine (de cytokinines). De precieze functie van deze bindingssite is onbekend, maar *Leguminosae* lectines spelen mogelijk een rol in groeiprocessen of opslag van plant hormonen. Op basis van foto-affiniteitslabeling van de adenine bindingssite in twee nauw verwante lectines (Lima boon lectine en PHA-E) werd voorgesteld dat adenine binding door PHA-L in de centrale holte gebeurt. Dit werd later bevestigd (zie verder). Een vergelijking van de monosaccharide bindingssite van PHA-L met die van andere *Leguminosae* lectines leidde tot de conclusie dat PHA-L geen monosaccharides bindt wegens het ontbreken van aromatische stacking tegen de suikerring.

Arceline is een getrunkeerd en dus inactief *Leguminosae* lectine uit de gewone boon dat nauw verwant is met PHA-L. Arceline komt alleen voor in bepaalde wilde boonvariëteiten en beschermt de plant tegen predatie door de "Mexican bean weevil" (*Zabrotes subfasciatus*). Het mechanisme achter de toxiciteit van arceline is echter onbekend.

De structuur van arceline-5 werd opgelost bij een resolutie van 2.7 ÅŻoals verwacht lijkt de architectuur van arceline-5 sterk op die van een *Leguminosae* lectine, met twee belangrijke verschillen. Ten eerste kristalliseerde arceline-5 als een monomeer, terwijl alle gekende *Leguminosae* lectines voorkomen als tetrameren of dimeren. Ten tweede mist arceline de geconserveerde metaalbindingssite, die een eveneens geconserveerde *cis*-peptide binding tussen een Ala en een Asp residu stabiliseert. Deze *cis*-peptide binding is van cruciaal belang voor de suikerbinding. Ver-

rassend genoeg bezit arceline-5 een Ala-Tyr *cis*-peptide binding op dezelfde plaats, ondanks de afwezigheid van de stabiliserende metaalionen. Inspectie van alle gekende arceline sequenties toont aan dat de Ala-Tyr of Ala-Phe sequentie steeds teruggevonden wordt. Aangezien niet X-Pro *cis*-peptide bindingen zeldzaam zijn en vaak een belangrijke functionele of katalytische rol spelen is het waarschijnlijk dat de Ala-Tyr/Phe *cis*-peptide binding belangrijk is voor de functie van de arcelines.

Het *Dolichos biflorus* zaad lectine (DBL) heeft een interessante specificiteit omdat het GalNAc goed bindt, terwijl het Gal slechts zeer zwak bindt. De meeste *Leguminosae* lectines binden Gal en GalNAc met een vergelijkbare affiniteit. Voorts is DBL, net als PHA-L, een adenine en plantenhormoon bindend tetrameer.

DBL werd gekristalliseerd in aanwezigheid van het bloodgroep A trisaccharide GalNAc(α1–3)[Fuc(α1–2)]Gal (2.8 Å), het Forssman disaccharide GalNAc(α1–3)GalNAc (2.6 Å) en adenine (2.65 Å). In de bloedgroep A trisaccharide structuur is alleen het GalNAc residu zichtbaar in de densiteit. Vergelijking met Gal-bindende *Leguminosae* lectines wijst uit dat DBL Gal bindt met lage affiniteit vanwege het ontbreken van aromatische stacking tegen de Gal ring : DBL heeft een alifatisch residu (Leu 127) op de plaats waar in andere *Leguminosae* lectines een aromaat aanwezig is. GalNAc bindt echter met een veel hogere affiniteit dan Gal omdat de N-acetyl groep compenseert voor het verlies van aromatische stacking door te interageren met Gly 103, Tyr 104 en Trp 132.

De niet-reducerende GalNAc van het Forssman disaccharide wordt op precies dezelfde wijze gebonden als de GalNAc van het bloedgroep A trisaccharide. De reducerende GalNAc van het disaccharide gaat geen additionele waterstofbruggen aan met het proteïne, maar creëert een hydrofobe cluster gevormd door Leu 127, Tyr 218, de hydrophobe patch van de niet reducerende GalNAc (C5 en C6) en de methyl groep van de reducerende GalNAc. De hogere affiniteit van DBL voor het Forssman disaccharide is dus uitsluitend te wijten aan hydrophobe interacties.

De quaternaire structuur van DBL is dezelfde als die van PHA-L, dat eveneens adenine bindt. In het geval van DBL kon de densiteit in het centrale kanaal van het tetrameer geïnterpreteerd worden als twee α-helixen die geklemd worden tussen de β-sheets van de twee kanonische dimeren. De adenine bindingsite bevindt zich in de buurt van een subeenheid interface en vormt vier potentiële adenine bindingssites, waarvan er slechts twee tegelijk kunnen ingenomen worden.

Chapter 7

Abbreviations used

- aAI α-amylase inhibitor (*Phaseolus vulgaris*)
- ANS 1,8-anilino-naphtalenesulfonic acid
- Arc5 Arcelin-5 (*Phaseolus vulgaris*)
- aSFP Acidic seminal fluid protein
- CC Correlation coefficient
- DB58 Horse gram (*Dolichos biflorus*) stem- and leaves lectin
- DBL Horse gram (*Dolichos biflorus*) seed lectin
- Con A Jack bean (*Canavalia ensiformis*) lectin
- Con B Concanavalin B
- EcorL West Indian coral tree (*Erythrina corallodendron*) lectin
- ER Endoplasmatic reticulum
- ERGIC-53 Type I membrane protein of the endoplasmatic reticulum-Golgi intermediate compartment
- Fuc *L*-Fucose
- Gal *D*-Galactose
- GalNAc *N*-Acetyl-*D*-Galactosamine
- Glc *D*-Glucose
- GlcNAc *N*-Acetyl-*D*-glucosamine
- GS4 *Griffonia simplicifolia* (*now Bandeirea simplicifolia*) isolectin 4
- LOL *Lathyrus ochrus* isolectins
- Man *D*-Mannose
- NCS Non crystallographic symmetry
- PHA-E Erythroagglutinating common bean (*Phaseolus vulgaris*) agglutinin

- PHA-L Leucoagglutinating common bean (*Phaseolus vulgaris*) agglutinin
- PNA Peanut (*Arachis hypogaea*) agglutinin
- PSP-I/PSP-II Porcine seminal plasma proteins I and II
- Rmsd Root mean square deviation
- TNS 2,6-toluidinyl-naphtalene-sulfonic acid
- SA Simulated annealing
- SBA Soybean (*Glycine max*) agglutinin
- VIP36 Vesicular integral membrane protein (36 kD)
- WBAI Basic winged bean (*Psophocarpus tetragonolobus*) agglutinin