

The Crystallographic Structure of Phytohemagglutinin-L*

(Received for publication, December 7, 1995, and in revised form, March 25, 1996)

Thomas W. Hamelryck[‡], Minh-Hoa Dao-Thi[‡], Freddy Poortmans[¶], Maarten J. Chrispeels^{||},
Lode Wyns[‡], and Remy Loris^{‡**}

From the [‡]Dienst Ultrastructuur, Vlaams Interuniversitair Instituut voor Biotechnologie, Vrije Universiteit Brussel, Paardenstraat 65, B-1640 Sint-Genesius-Rode, Belgium, the [¶]Vlaamse Instelling voor Technologisch Onderzoek-VITO, Boeretang 200, B-2400 Mol, Belgium, and the ^{||}Department of Biology, University of California, San Diego, La Jolla, California 92093-0116

The structure of phytohemagglutinin-L (PHA-L), a leucoagglutinating seed lectin from *Phaseolus vulgaris*, has been solved with molecular replacement using the coordinates of lentil lectin as model, and refined at a resolution of 2.8 Å. The final R-factor of the structure is 20.0%. The quaternary structure of the PHA-L tetramer differs from the structures of the concanavalin A and peanut lectin tetramers, but resembles the structure of the soybean agglutinin tetramer. PHA-L consists of two canonical legume lectin dimers that pack together through the formation of a close contact between two β -strands. Of the two covalently bound oligosaccharides per monomer, only one GlcNAc residue per monomer is visible in the electron density. In this article we describe the structure of PHA-L, and we discuss the putative position of the high affinity adenine-binding site present in a number of legume lectins. A comparison with transthyretin, a protein that shows a remarkable resemblance to PHA-L, gives further ground to our proposal.

Lectins are proteins that bind carbohydrates in a reversible and specific manner, and often have hemagglutinating properties. At present, the legume lectins undoubtedly form the most extensively studied group of lectins (Sharon and Lis, 1991). Over 70 different seed lectins have been identified in various *Leguminosae* species. The function of these lectins *in vivo* remains unsure, but defense against predation (Chrispeels and Raikhel, 1991) and interaction with symbionts (Diaz *et al.*, 1989) have been proposed. The x-ray structures of eight different legume lectins have been solved and refined: pea lectin (PSL, Einspahr *et al.* (1986)), lentil lectin (LCL, Loris *et al.* (1993)), *Lathyrus ochrus* isolectins I and II (LOL I, Bourne *et al.* (1990); LOL II, Bourne *et al.* (1994)), *Griffonia simplicifolia* lectin IV (GS4, Delbaere *et al.* (1990)), *Erythrina corallodendron* lectin (EcorL, Shaanan *et al.* (1991)), concanavalin A (Con A, Becker *et al.* (1975)), peanut agglutinin (PNA, Banerjee *et al.* (1994)) and, recently, soybean agglutinin (SBA, Dessen *et al.* (1995)). The seeds of the common bean contain a protein frac-

tion with sugar binding and hemagglutinating properties, called phytohemagglutinin (PHA).¹ This fraction consists of five different tetramers, built out of two polypeptides (L and E) in all possible combinations (L₄, L₃E, L₂E₂, LE₃, E₄). The E-type and L-type subunits are, respectively, responsible for the erythroagglutinating and leucoagglutinating properties of the PHA fraction. The two polypeptides 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. 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 (Moreno and Chrispeels, 1989). Arcelin and α -amylase inhibitor can be considered as truncated forms of PHA, in which, respectively, one and two loops that play a sugar binding role are missing, abolishing the sugar binding properties. Both arcelin (Osborn *et al.*, 1988) and α -amylase inhibitor (Shade *et al.*, 1994) protect bean seed against predation by pests although the precise origin of the toxicity of arcelin is as yet unknown. In this paper we present the crystal structure of PHA-L, the L₄ tetramer. 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 and Chrispeels, 1986). The minimal structural unit for high-affinity binding by PHA-L is the pentasaccharide Gal β 1 \rightarrow 4GlcNAc β 1 \rightarrow 2[Gal β 1-4GlcNAc β 1-6]Man, which is found in tetra- and tri-antennary complex type oligosaccharides of mammalian origin (Hammerström *et al.*, 1982). In addition to the sugar binding properties of the legume lectins, a number of legume lectins also bind adenine and related ligands. High affinity binding sites for adenine and its derivatives have been found for PHA-E, *Dolichos biflorus* seed lectin (DBL), *D. biflorus* stem and leaves lectin (DB58), soybean agglutinin, and *Phaseolus lunatus* lectin (LBL) (Roberts and Goldstein, 1982; Roberts and Goldstein, 1983a; Gegg *et al.*, 1992). In this article, we suggest the possible location of this site, based on photoaffinity labeling of the adenine site in PHA-E and LBL (Maliarik and Goldstein, 1988) and the common quaternary structure of PHA-L and SBA.

* This work was supported by grants from the Vlaams Actieprogramma Biotechnologie and the Vlaams Interuniversitair Instituut voor Biotechnologie projects of the Flemish government. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Supported by a grant from the Vlaams Instituut voor de Bevordering van het Wetenschappelijk-Technologisch Onderzoek in de Industrie (IWT). To whom correspondence should be addressed: Dienst Ultrastructuur, Vlaams Interuniversitair Instituut voor Biotechnologie, Vrije Universiteit Brussel, Paardenstraat 65, B-1640 Sint-Genesius-Rode, Belgium. Tel.: 32-2-3590282; Fax: 32-2-3590289; E-mail: thamelry@vub.ac.be.

** Senior research assistant of the Nationaal Fonds voor Wetenschappelijk Onderzoek.

¹ The abbreviations used are: PHA-L, phytohemagglutinin-L; PHA-E, phytohemagglutinin-E; PSL, pea lectin; LCL, lentil lectin; SBA, soybean agglutinin; PNA, peanut agglutinin; GS4, *Griffonia simplicifolia* isolectin 4; ConA, concanavalin A; EcorL, *Erythrina corallodendron* lectin; LBL, Lima bean lectin; DB58, *Dolichos biflorus* stem and leaves lectin; DBL, *Dolichos biflorus* seed lectin; ANS, 1,8-anilinonaphthalene-sulfonic acid; TNS, 2,6-toluidinylnaphthalenesulfonic acid; r.m.s., root mean square.

TABLE I
Data collection statistics of the native PHA crystal

Resolution	Number of observed reflections	Number of unique reflections	Completeness	Multiplicity	R_{merge}^a
Å			%		
6.78–10.00	1,632	649	96.0	1.7	0.051
5.46–6.78	4,309	1,320	97.6	2.7	0.080
4.70–5.46	7,671	2,019	96.7	3.7	0.079
4.19–4.70	8,552	2,287	96.8	3.6	0.090
3.81–4.19	6,925	2,144	94.7	2.9	0.099
3.52–3.81	5,158	1,893	91.0	2.2	0.081
3.29–3.52	5,265	1,946	90.2	2.2	0.111
3.10–3.29	5,066	1,881	87.6	2.1	0.149
2.94–3.10	5,036	1,871	85.1	2.0	0.209
2.80–2.94	4,923	1,832	81.8	2.0	0.293
Total	61,545	24,850	88.5	2.5	0.097

$$^a R_{\text{merge}} = \sum_i \sum_j (|I_{ij} - \langle I_j \rangle|) / \sum_i I_{ij}$$

EXPERIMENTAL PROCEDURES

Crystallization and Data Collection—Crystallization of PHA-L has been described elsewhere (Dao-Thi *et al.*, 1996). 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 μ l of bottom solution and 5 μ l of 5 mg/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 $\beta = 93.7^\circ$. The asymmetric unit contains one complete PHA-L tetramer. Data collection was performed on a single crystal at room temperature with an Enraf-Nonius FAST area detector, using CuK α radiation generated by a rotating anode x-ray generator (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 I and II.

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 by 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 data base) 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, 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, $t = 0.0001$ 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 $^{-2}$, $\sigma_{\text{ncs}} = 1.0$ Å 2). After each refinement step the quality of the model was checked using PROCHECK (Laskowski *et al.*, 1993). We 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, 1996), 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 data base was used to compare PHA-L with SBA. Accessible surface areas were calculated with the GRASP program (Nicholls *et al.*, 1991).

RESULTS AND DISCUSSION

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 r.m.s. 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 mature monomer, based on the sequence of the PHA-L encoding gene (Hoffman and Donaldson, 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 and Chrispeels, 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 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. As is the case with the other solved legume lectin structures, the subunit structure in PHA-L consists of a flat, six-stranded β -sheet, and a curved, seven-stranded β -sheet.

Quaternary Structure—PHA-L is a tetramer with approximate dimensions 40 Å \times 60 Å \times 80 Å, consisting of four identical subunits (Fig. 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 12-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). In the tetramer, the two curved 12-strand β -sheets face each other, creating a large channel between them. The same tetrameric organization is also found in SBA (Dessen *et al.*, 1995). As most tetramers (Miller, 1989), the PHA-L tetramer has internal 222 symmetry. At present, the x-ray structures of three other legume lectin tetramers (PNA, SBA, and ConA) and of three different types of dimers (GS4, EcorL and the "canonical dimer," as represented by lentil lectin, the two *L. ochrus* isolectins, and pea lectin) 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 ConA, SBA, and PHA-L, this protein does not possess 222 symmetry. Although SBA and PHA-L on the one hand and ConA on the other hand both possess 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. 1). ConA 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 toward the inside of

TABLE II
Overview of the statistics of data collection and refinement

Space group	C2
Unit cell	$a = 106.3 \text{ \AA}$ $b = 121.2 \text{ \AA}$ $c = 90.8 \text{ \AA}$ $\beta = 93.7^\circ$
Resolution	2.8 \AA
Number of observations	61,545
Number of unique reflections	24,850
Multiplicity	2.5
Completeness (2.80 \AA–2.94 \AA, last shell)	88.5% (81.8%)
R_{merge} (2.80 \AA–2.94 \AA, last shell)	9.7% (29.3%)
Final R -value (10.0 \AA–2.8 \AA)	20.0%
A posteriori free R -value (10.0 \AA–2.8 \AA)	22.9%
Number of protein atoms	3,838
Number of solvent atoms	16
Number of sugar atoms	46
Number of metal ions	4 Mn^{2+} , 4 Ca^{2+}
R.m.s. on bond lengths	0.017 \AA
R.m.s. on bond angles	2.01 $^\circ$
R.m.s. on dihedrals	28.16 $^\circ$
R.m.s. on impropers	1.71 $^\circ$
Distribution of non-Gly non-Pro ϕ - ψ 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

$$R = \sum(F_o - F_c) / \sum F_o; R_{\text{merge}} = \sum_i \sum_j \langle |I_i - I_{ij}| / \sum_j I_{ij} \rangle$$



FIG. 1. **Comparison between ConA and PHA-L.** The PHA-L tetramer is shown on the left, the ConA tetramer is shown on the right. The left dimers in both tetramers have the same orientation to emphasize the difference in dimer-dimer packing between PHA-L and ConA. The central channel running between the two dimers in PHA-L is clearly visible. All figures were made using MOLSCRIPT (Kraulis, 1991).

the monomer and toward 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 III). The fact that the protruding side chains are small (Ile and Ser), allows them to be intercalated in a “zipper-like” fashion (Fig. 2). This 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). This kind of architecture 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 silk β -sandwich to be closer to each other than in the default packing: the backbone atoms are only 6 \AA apart in pyruvoyl-dependent histidine decarboxylase, as opposed to 10 \AA in a classic β -sheet. The backbone atoms of the two interacting β -strands in PHA-L are 6.5 \AA apart. Eight hydrogen bonds are made in the interface between the two dimers, involving two serine side chain hydroxyl groups (Ser-186 and Ser-190) and a

lysine main chain oxygen atom (Lys-184) in each monomer (see Table IV). This lysine 184 residue 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, and Thr-192). The way in which the two dimers pack in PHA-L creates a large channel in the center of the molecule: the two 12-strand β -sheets facing each other are approximately 18 \AA apart in the center. Similar channels are found in tetramers where the monomers, as in PHA-L and ConA, only pairwise interact (Miller, 1989).

The Metal Binding Site—All legume lectins possess two bound metal ions (one calcium 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: 1 His, 1 Glu, 1 Asn, 2 Asp, and 1 hydrophobic residue (Phe in LCL, PSL, LOL I and II, EcorL, SBA; Tyr in ConA 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. 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, see above), that is involved in hydrophobic contacts with the bound sugar. The Ca^{2+} ion interacts via a water molecule with an oxygen atom of the side chain of Asp-86, and thus stabilizes the conserved Ala-85, Asp-86 cis-peptide bond, that is present in all solved legume lectin structures.

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 2 Ser (Ser-187 and Ser-191 in SBA) and 1 Ile residue (Ile-189 in SBA). The only substantial difference between the two interfaces is the con-

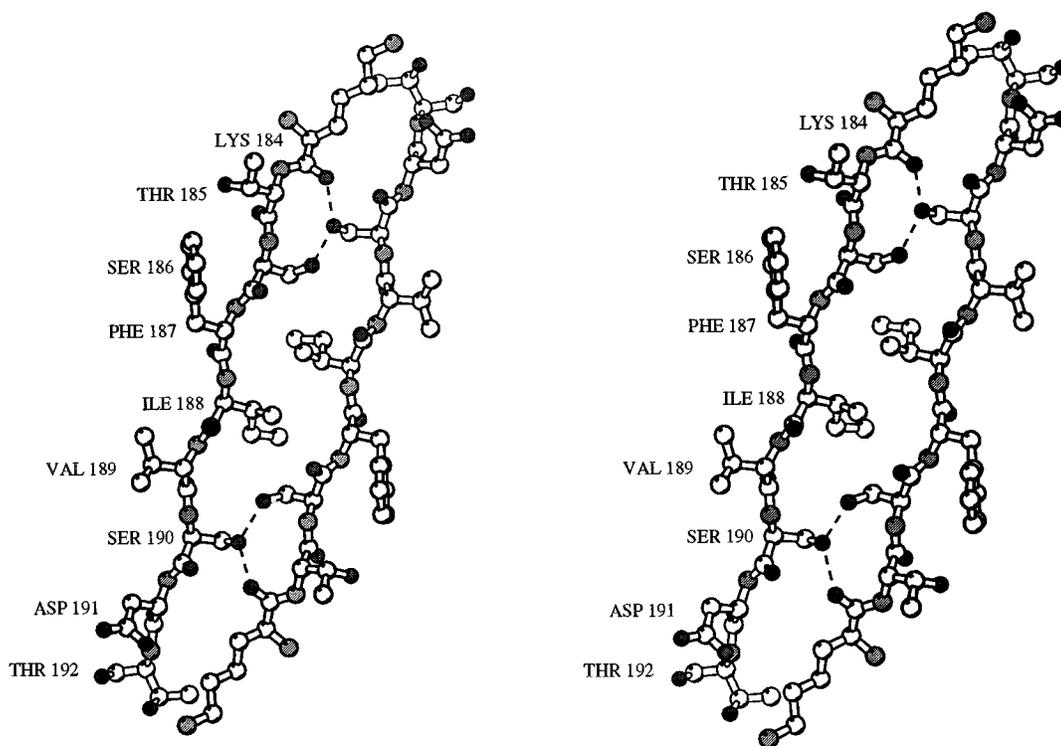


FIG. 2. Stereoview of the two strands that form the 2-fold symmetrical interface between the two canonical dimers in PHA-L. The 2-fold axis is approximately positioned in the center of the figure. The side chains of residues Ser-186, Ile-188, and Ser-190 intercalate, as is clearly visible in the figure. The four hydrogen bonds (Lys-184A O-Ser-190C OG, Ser-186A OG-Ser-190C OG, Ser-190A OG-Lys-184C O, and Ser-190A OG-Ser-186C OG) across the interface are visualized as *dashed lines*. The hydrophobic part of the side chain of Lys-184 makes extensive van der Waals contacts with Ser-190, Asp-191, and Thr-192 from the other strand. Carbon atoms are white, nitrogen atoms are light gray, oxygen atoms are dark gray.

TABLE III

van der Waals contacts between the two canonical dimers in PHA-L for the A-C dimer interface. Atom distances below 4 Å were 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

servative 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 the side chain also forms an additional salt bridge with the Asp-192 residue. The structure of the SBA dimer-dimer interface is shown in Fig. 3.

It has been suggested that in SBA, a C-truncated subunit (240 AA) and a non-truncated subunit (253 AA) 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 not investigated for PHA-L. The intact PHA-E subunit consists of 254 AA, the truncated subunit consists of 244 AA. In

TABLE IV

Hydrogen bonds between the two canonical dimers in the PHA-L tetramer

Dimer 1	Dimer 2	Distance
		Å
Lys-184A O	Ser-190C OG	2.94
Ser-186A OG	Ser-190C OG	2.70
Ser-190A OG	Lys-184C O	3.02
Ser-190A OG	Ser-186C OG	2.57
Lys-184B O	Ser-190D OG	2.81
Ser-186B OG	Ser-190D OG	2.83
Ser-190B OG	Lys-184D O	3.06
Ser-190B OG	Ser-186C OG	2.76

PHA-L, no electron density is observed beyond 233, due to either C-terminal truncation or chain flexibility, or possibly a combination of both. In any case, 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 Å² of the protein surface in PHA-L, comparable to 1200 Å² for SBA.

Putative Position of the Adenine Binding Site—In addition to the sugar binding site, a number of legume lectins possess hydrophobic binding sites that can be divided into three groups. The first site is adjacent to the sugar binding site and is responsible for the 10-fold higher affinity of these lectins for monosaccharides bearing hydrophobic substituents, as compared to the nonsubstituted monosaccharides. The second site is positioned at a distance of approximately 30 Å from the sugar binding site and has a low affinity (10² M⁻¹) for hydrophobic ligands like indoleacetic acid, TNS and ANS. Finally, the third site is a high affinity (10⁶ M⁻¹) binding site for adenine and its

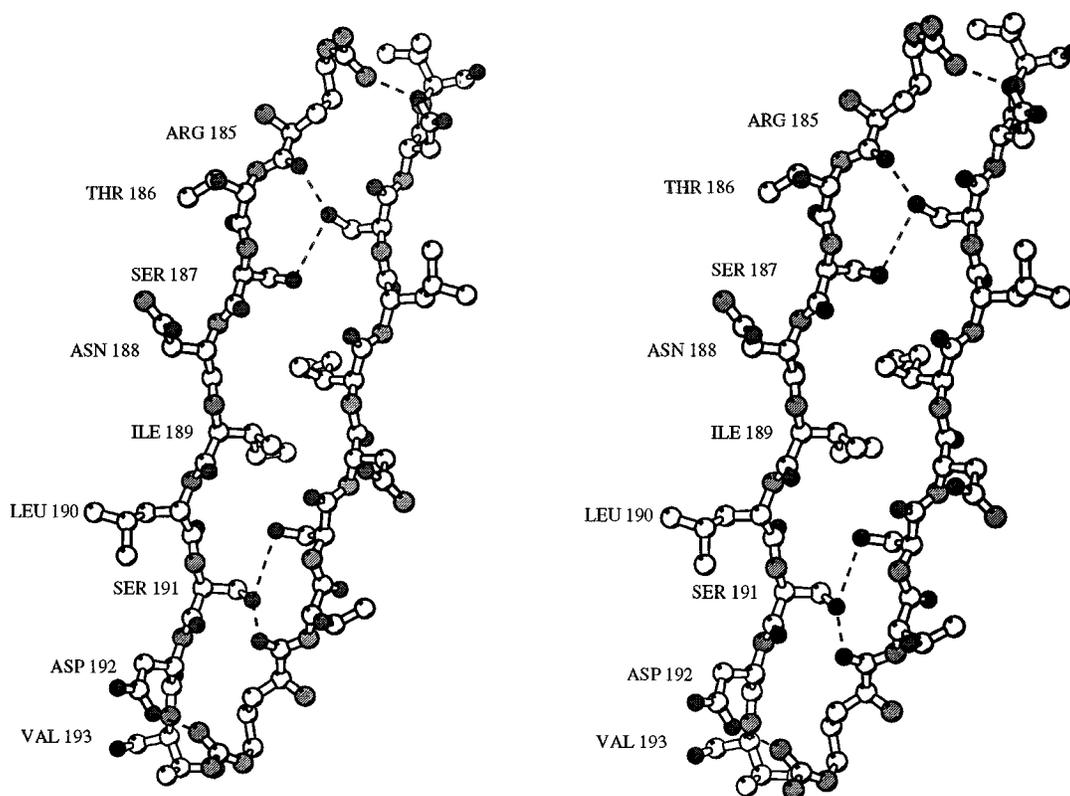


FIG. 3. A stereo figure of the dimer-dimer interface in the SBA tetramer. The interface is similar to the dimer-dimer interface in PHA-L. The 2-fold axis is approximately positioned in the center of the figure. As in PHA-L, the side chains of 2 Ser (Ser-191 and -187) and 1 Ile residue (Ile-189) intercalate. The main difference with PHA-L is the substitution of Lys-184 in PHA-L by Arg-185 in SBA. The Arg-185 is involved in an additional hydrogen bond with Asp-192. The six hydrogen bonds (Arg-185A O-Ser-191C OG, Arg-185A NH1-Asp-192C OD1, Ser-187A OG-Ser-191C OG, Ser-191A OG-Arg-185C O, Ser-191A OG-Ser-187C OG, Asp-192A OD1-Arg-185C NH1) across the interface are visualized as dashed lines.

N-6 derivatives, including a number of adenine-derived plant hormones. This site also binds TNS, but not ANS. High affinity binding of adenine has been shown for SBA, DBL, DB58, PHA-E, and LBL. 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, 1983b), but more recent research yielded two adenine binding sites for the DBL tetramer (Gegg *et al.*, 1992), suggesting a symmetric binding site at the interface between two subunits, as pointed out by Gegg *et al.* (1992). The sequences of the adenine binding sites in LBL and PHA-E have been determined by photoaffinity labeling (Maliarik and Goldstein, 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 and Goldstein, 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% homology) 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 certain legume lectins. In the PHA-L structure, this sequence is found in the third strand of the flat 6-strand β -sheet, from residue 163 to 167 (Fig. 4). This β -sheet is located in the central channel present in PHA-L. If this region is indeed involved in adenine binding, then localization 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).

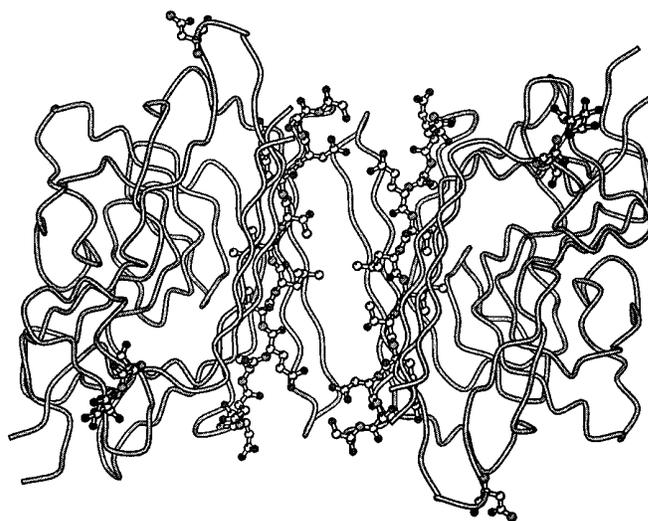


FIG. 4. A view from the center of the molecule to the dimer-dimer interface on the putative adenine binding site between dimers A and C. The strand that contains the photoaffinity labeled residues in both monomers is shown as a ball-and-stick representation. The two glycosylated residues per monomer (Asn-12 and Asn-60) are also shown in ball-and-stick representation, together with the GlcNAc residue bound to Asn-12.

The organization of the PHA-L tetramer is such that the labeled 5-residue sequences of two monomers are close together, and are related by the 2-fold axis along the length of the molecule, thus creating a binding site with a 2-fold symmetry. Apparently, the adenine binding site is located near the inter-

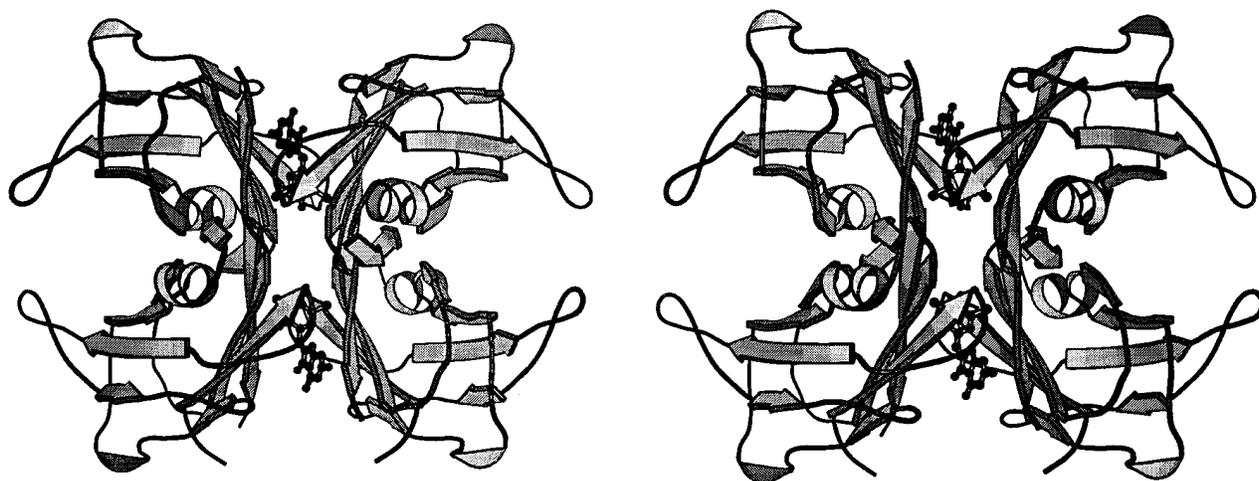


FIG. 5. Stereo figure of the transthyretin tetramer complexed with 3',5'-dibromo-2',4,4',6-tetrahydroxyaurone, a flavone derivative. The ligand is bound in the central hole that runs through the molecule, through interactions with side chains of the residues that make up the flanking β -strands (mainly Ser, Thr, Ala, and Leu). The binding site possesses 2-fold symmetry. The central hole is approximately 10 Å wide.

face 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, far too great to cause direct steric hindrance between two bound adenine molecules. Possibly one bound adenine molecule impairs binding of a second one by an allosteric mechanism, resulting in only one observed adenine binding site. In this respect, it is useful to mention that the binding of certain ligands in the adenine binding site can enhance or diminish the affinity of the hydrophobic binding site for ANS (Roberts and Goldstein, 1983a).

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.*, 1978) (Fig. 5). 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 2-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, and 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.

Concluding Remarks—In this paper we described the structure of PHA-L. PHA-L is a tetrameric protein, in which two canonical legume lectin dimers pack together through the formation of a close contact between the two outmost β -strands. This way of packing creates a large channel in the center of the protein, lined mainly with Ser, Thr, and small, apolar residues (Val, Ala, Ile, and Leu). We suggest that this channel forms the adenine binding site, based on photoaffinity labeling (Maliarik and Goldstein, 1988) and comparison with transthyretin. The

parallels between our model of PHA-L and transthyretin are striking: (i) both consist of similar dimers, (ii) the packing of both dimers in the tetramer is similar, (iii) 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 the two β -sheets, (iv) both transthyretin and PHA-L bind similar ligands (ANS and TNS). Our results indicate that PHA-L and transthyretin apparently converged toward a similar framework for binding small ligands with aromatic groups.

Acknowledgment—We thank Maria Vanderveken for excellent technical assistance.

REFERENCES

- Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Suroliya, A., and Vijayan, M. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 227–231
- Becker, J. W., Reeke, G. N., Jr., Wang, J. L., Cunningham, B. A., and Edelman, G. M. (1975) *J. Biol. Chem.* **250**, 1513–1524
- Blake, C. C. F., and Oatley, S. J. (1977) *Nature* **268**, 115–120
- Blake, C. C. F., Geisow, M. J., Oatley, S. J., Rerat, B., and Rerat, C. (1978) *J. Mol. Biol.* **121**, 339–356
- Bourne, Y., Roussel, A., Frey, M., Rougé, P., Fontecilla-Camps, J.-C., and Cambillau, C. (1990) *Proteins* **7**, 365–376
- Bourne, Y., Mazurier, J., Legrand, D., Rougé, P., Montreuil, J., Spik, G., and 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 Crystallogr.* **D49**, 24–36
- Casset, F., Hamelryck, T., Loris, R., Brisson, J., Tellier, C., Dao-Thi, M., Wyns, L., Poortmans, F., Pérez, S., and Imberty, A. (1995) *J. Biol. Chem.* **270**, 25619–25628
- Chothia, C., and Janin, J. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 4146–4150
- Chrispeels, M. J., and Raikhel, N. V. (1991) *Plant Cell* **3**, 1–9
- Ciszak, E., Cody, V., and Luft, R. L. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 6644–6648
- Dao-Thi, M.-H., Hamelryck, T. W., Poortmans, F., Voelker, T. A., Chrispeels, M. J., and 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., and Lemieux, R. (1990) *Can. J. Chem.* **68**, 1116–1121
- Diaz, C., Melchers, L. S., Hooykaas, P. J. J., Lugtenberg, B. J. J., and Kijne, J. W. (1989) *Nature* **338**, 579–581
- Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F., and Sacchettini, J. C. (1995) *Biochemistry* **34**, 4933–4942
- Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E., and Suddath, F. L. (1986) *J. Biol. Chem.* **261**, 16518–16527
- Ferguson, R. N., Edelhofer, H., Saroff, H. A., and Robbins, J. (1975) *Biochemistry* **14**, 282–289
- Gallagher, T., Rozwarski, D. A., Ernst, S. R., and Hackert, M. L. (1993) *J. Mol. Biol.* **230**, 516–528
- Gegg, C. V., Roberts, D. D., Segel, I. H., and Etzler, M. E. (1992) *Biochemistry* **31**, 6938–6942
- Goossens, A., Geremia, R., Bauw, G., Van Montagu, M., and Angenon, G. (1994) *Eur. J. Biochem.* **225**, 787–795
- Hammerström, S., Hammerström, M., Sundblad, G., Arnarp, J., and Lönngren, J. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 1611–1615

- Hartweck, L. M., Vogelzang, R. D., and Osborn, T. C. (1991) *Plant Physiol.* **97**, 204–211
- Hoffman, L. M., and Donaldson, D. D. (1985) *EMBO J.* **4**, 883–889
- Jordan, E. T., and Goldstein, I. J. (1994) *J. Biol. Chem.* **269**, 7674–7681
- Kraulis, P. J. (1991) *J. Appl. Cryst.* **24**, 946–950
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst.* **26**, 283–291
- Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R., and Wyns, L. (1993) *Biochemistry* **32**, 8772–8781
- Loris, R., Stas, P. P. G., and Wyns, L. (1994) *J. Biol. Chem.* **269**, 26722–26733
- Maliarik, M. J., and Goldstein, I. J. (1988) *J. Biol. Chem.* **263**, 11274–11279
- 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 Eng.* **3**, 77–83
- Moreno, J., and Chrispeels, M. J. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7885–7889
- Murzin, A. G. (1994) *Curr. Opin. Struct. Biol.* **4**, 441–449
- Navaza, J. (1994) *Acta Cryst. A* **50**, 157–163
- Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins* **11**, 281–296
- Osborn, T. C., Alexander, D. C., Sun, S. S. M., Cardona, C., and Bliss, F. A. (1988) *Science* **240**, 207–210
- Pflugrath, J. W., and Messerschmidt, A. (1989) *MADNESS manual of FAST diffractometer*, Enraf-Nonius, Delft, The Netherlands
- Roberts, D. D., and Goldstein, I. J. (1982) *J. Biol. Chem.* **257**, 11274–11277
- Roberts, D. D., and Goldstein, I. J. (1983a) *J. Biol. Chem.* **258**, 13820–13824
- Roberts, D. D., and Goldstein, I. J. (1983b) *Chemical Taxonomy, Molecular Biology, and Function of Plant Lectins* (Goldstein, I. J., and Etzler, M. E., eds) pp. 131–141, Alan R. Liss, New York
- Romero Andreas, J., Yandell, B. S., and Bliss, F. A. (1986) *Theor. Appl. Genet.* **72**, 123–128
- Shaanan, B., Lis, H., and 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., and Chrispeels, M. J. (1994) *Bio/technology* **12**, 793–796
- Sharon, N., and Lis, H. (1990) *FASEB J.* **4**, 3198–3208
- Sturm, A., and Chrispeels, M. J. (1986) *Plant Physiol.* **80**, 320–322
- Wojtczak, A., Luft, J., and Cody, V. (1992) *J. Biol. Chem.* **267**, 353–357
- Young, N. M., Watson, D. C., Yaguchi, M., Adar, R., Arango, R., Rodriguez-Arango, E., Sharon, N., Blay, P. K. S., and Thibault, P. (1995) *J. Biol. Chem.* **270**, 2563–2570