THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 51, Issue of December 20, pp. 32796–32802, 1996
Printed in U.S.A.

Crystal Structure of Arcelin-5, a Lectin-like Defense Protein from *Phaseolus vulgaris**

(Received for publication, July 25, 1996, and in revised form, September 17, 1996)

Thomas W. Hamelryck[‡][§][¶], Freddy Poortmans[‖], Alain Goossens[**][§], Geert Angenon[**], Mark Van Montagu[**], Lode Wyns[‡], and Remy Loris[‡] [‡‡]

From [‡]Laboratorium voor Ultrastructuur, Vlaams Interuniversitair Instituut voor Biotechnologie, Vrije Universiteit Brussel, Paardenstraat 65, B-1640 Sint-Genesius-Rode, Belgium, [‖]Vlaamse Instelling voor Technologisch Onderzoek (VITO), Boeretang 200, B-2400 Mol, Belgium, and [**]Laboratorium voor Genetica, Vlaams Interuniversitair Instituut voor Biotechnologie, Universiteit Gent, K. L. Ledeganckstraat 35, B-9000 Gent, Belgium

In the seeds of the legume plants, a class of sugar-binding proteins with high structural and sequential identity is found, generally called the legume lectins. The seeds of the common bean (*Phaseolus vulgaris*) contain, besides two such lectins, a lectin-like defense protein called arcelin, in which one sugar binding loop is absent. Here we report the crystal structure of arcelin-5 (Arc5), one of the electrophoretic variants of arcelin, solved at a resolution of 2.7 Å. The *R* factor of the refined structure is 20.6%, and the free *R* factor is 27.1%. The main difference between Arc5 and the legume lectins is the absence of the metal binding loop. The bound metals are necessary for the sugar binding capabilities of the legume lectins and stabilize an Ala-Asp *cis*-peptide bond. Surprisingly, despite the absence of the metal binding site in Arc5, this *cis*-peptide bond found in all legume lectin structures is still present, although the Asp residue has been replaced by a Tyr residue. Despite the high identity between the different legume lectin sequences, they show a broad range of quaternary structures. The structures of three different dimers and three different tetramers have been solved. Arc5 crystallized as a monomer, bringing the number of known quaternary structures to seven.

Lectins, a class of sugar-binding proteins without enzymatic activity toward the bound sugar, are ubiquitous in nature and have been found in viruses, bacteria, plants, and animals. The most extensively studied class of lectins is undoubtedly the legume lectin class, found in the seeds of the legume plants (1). At present, the crystal structures of 10 different legume lectins have been solved and refined: pea lectin (2), lentil lectin (3), *Lathyrus ochrus* isolectins I (4) and II (5), *Griffonia simplicifolia* lectin IV (GS4)[1] (6), *Erythrina corallodendron* lectin (EcorL) (7), concanavalin A (8), peanut agglutinin (9), soybean agglutinin (10), and recently phytohemagglutinin-L (11). Despite the abundance of structural and biochemical information, the precise function of the legume lectins in the plant is still the subject of much debate. The two most likely functions are protection against predators (12) and interaction with the *Rhizobium* symbionts responsible for nitrogen fixation (13). In the seeds of the common bean, *Phaseolus vulgaris*, two lectins are found, called phytohemagglutinin-L (PHA-L) and phytohemagglutinin-E. The crystal structure of the former has recently been determined at a resolution of 2.8 Å (11). In addition, two polypeptides with highly related amino acid sequences are present in the seeds, namely arcelin and α-amylase inhibitor. In these so-called lectin-like proteins, respectively one and two loops essential for the sugar-binding capabilities of the legume lectins are deleted. The α-amylase inhibitor 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, the α-amylase inhibitor, and arcelin are defense proteins that protect the bean against predators (12), although the precise 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. Arcelin-5 (15) has been reported to consist of a mixture of two major protein fractions, termed arcelin-5a (Arc5a, 32.2 kDa) and arcelin-5b (Arc5b, 31.5 kDa), and a third, minor fraction, termed arcelin-5c (Arc5c, 30.8 kDa). Arc5b and Arc5a contain one and two glycans, respectively, while Arc5c is not glycosylated. Two different arcelin-5 cDNA sequences were reported (15), called *arc5-I* and *arc5-II*. They encode two polypeptides of 240 amino acids (26.8 and 27.0 kDa) 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. Arcelins are thought to provide resistance against the bean bruchid pest *Zabrotus subfasciatus* (16). Among the arcelin variants, Arc1 and Arc5 appear to be the most promising in conferring insect resistance (33). Here we present the crystal structure of arcelin-5 solved at a resolution of 2.7 Å.

* This work was supported in part by grants from the Vlaams Actieprogramma voor Biotechnologie (VLAB), the Vlaams Interuniversitair Instituut voor Biotechnologie (VIB), and the Algemeen Bestuur van de Ontwikkelingssamenwerking (ABOS). 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.

The atomic coordinates and structure factors for arcelin-5a (1IOA and R1IOASF) have been deposited in the Protein Data Bank, Brookhaven National Laboratory, Upton, NY.

[§] Recipient of a grant from the Vlaams Instituut voor de Bevordering van het Wetenschappelijk-technologisch Onderzoek in de Industrie (IWT).

[¶] To whom correspondence should be addressed. Tel.: 32-2-3590209; Fax: 32-2-3590289; E-mail: thamelry@vub.ac.be.

[‡‡] A senior research assistant of the Nationaal Fonds voor Wetenschappelijk Onderzoek (NFWO).

[1] The abbreviations used are: GS4, *G. simplicifolia* isolectin IV; NCS, noncrystallographic symmetry; PHA-L, phytohemagglutinin-L; SA, simulated annealing; ConB, concanavalin B; EcorL, *E. corallodendron* seed lectin.

TABLE I
Data collection statistics of the arcelin-5 crystal

Resolution	Observed reflections	Unique reflections	Completeness	Multiplicity	R_{merge}[a]
			%		%
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	49,089	16,722	95.9	2.9	8.6

[a] $R_{merge} = \sum_i\sum_j(I_{ij} - \langle Ij \rangle)/\sum_i I_{ij}$.

EXPERIMENTAL PROCEDURES

Crystallization and Data Collection—Arcelin-5 was purified from seeds of the wild *Phaseolus vulgaris* (strain G02771) by chromatofocusing as described in Ref. 15. Protein solutions were made starting from the lyophilized protein. Suitable crystallization conditions were screened using the hanging drop vapor diffusion method: 5 μl of 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 $(NH_4)_2SO_4$, sodium 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 microseeding 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 the aforementioned composition prior to equilibration. This resulted in a number of thin, plate-shaped crystals. The 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 copper Kα radiation generated by a rotating anode x-ray generator, operated at 40 kV and 90 mA. Autoindexing was done with the MADNESS software packet (17). The crystal was monoclinic and belonged to space group $P2_1$. The data were subdivided in bins of 3° and scaled and merged with the CCP4 programs ROTAVATA and AGROVATA (18). The relevant data collection details are given in Table I.

Structure Solution and Refinement—The structure of Arc5 was solved using molecular replacement with the program AMORE (19). The coordinates of PHA-L (11) 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 (15) and the identity between Arc5 and PHA-L is high (55%), molecular replacement was first attempted, however unsuccessfully, 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 is highly homologous to the legume lectins, the Arc5 sequence was submitted to the Swiss-Model service (20) 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) (21) on a Cray J916/8–1024 supercomputer. Positional refinement was done with the simulated annealing protocol (SA) (4000 K, t = 0.01 ps). In 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 (14), restrained noncrystallographic 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 interpret electron densities. Model visualization and building were done using the program O (version 5.10.2) (35) on a Silicon Graphics INDIGO^2 XZ workstation. The stereochemistry of the structure was monitored throughout the refinement with the program PROCHECK (22). Further details about the structure are given in Table II.

The coordinates of GS4, EcorL, and PHA-L, used for comparison with Arc5, are present in the Brookhaven Protein Data Bank 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 (23). The figures were made with the programs MOLSCRIPT (32) and O (35).

Gel Filtration—The molecular mass of the native arcelin-5 protein was determined by gel filtration through a Superdex 200 HR (Pharmacia Biotech Inc.) fast protein liquid chromatography column (1.0 × 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) was used as the equilibration and elution buffer. A gel filtration standard (Bio-Rad) with molecular weight markers ranging from 1.35 to 670 kDa was used.

RESULTS AND DISCUSSION

Overall Structure—The R factor of the final structure for all 16,722 reflections between 2.7 and 10.0 Å is 20.6%, and the free R factor is 27.1%. The overall quality of the electron density is quite high.

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. 1). 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

TABLE II
Overview of the statistics of data collection and refinement

Parameter	Value
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 (10.0–2.7 Å)[a]	20.6%
Final free R value (10.0–2.7 Å)	27.1%
Number of protein atoms	4554
Number of solvent atoms	0
Number of glycan atoms	76
r.m.s. on bond lengths[b]	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

[a] $R=\sum|F_o-F_c|/\sum|F_o|$; $R_{merge}=\sum_i\sum_j\langle I_i\rangle-I_{ij}/\sum_i\sum_j I_{ij}$
[b] r.m.s., root mean square.

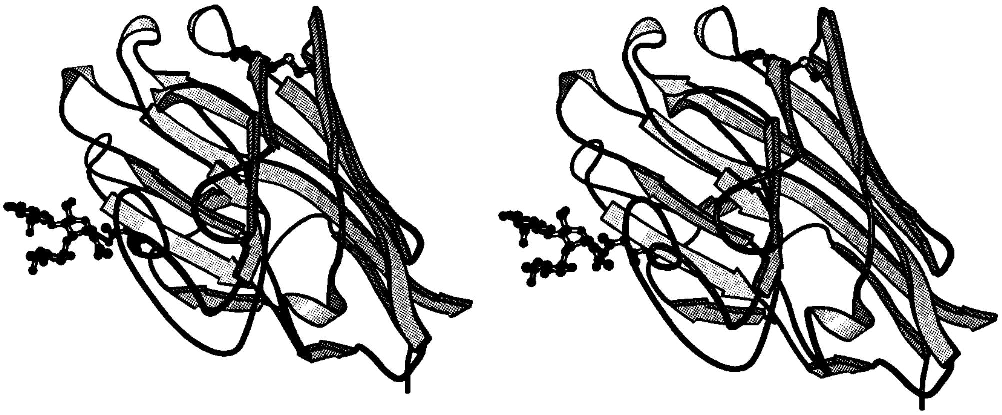

FIG. 1. **An overall view of the Arc5 monomer.** The glycan attached to Asn[22] (GlcNAcβ1→4(Fucα1→3)GlcNAc) and the disulfide bridge between residues 146 and 182 are shown as a *ball-and-stick model*. This cystine links the seventh β-strand of the front sheet and the sixth strand of the back sheet together.

details, see below). A fourth sugar residue displayed only weak electron density and was subsequently 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 with low B factors in both monomers present in the asymmetric unit, 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 root mean square difference between the positions of the superimposable C-α 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. 2). 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.

The Quaternary Structure of Arc5—Although Arc5 has been reported to be a dimer in solution (15), 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 mass of 25 or 77 kDa, respectively. The sample used for crystallization consisted of purified protein that had first been lyophilized and then resuspended in water. In this sample only the monomer form was present. This was also the case when the lyophilized protein was resuspended in phosphate-buffered saline or in acetate buffer. In a crude extract (partially purified albumins (15)) and in a purified fraction that had never been lyophilized 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 (concanavalin A, peanut agglutinin, and PHA-L/soybean agglutinin) and three types of dimers (EcorL, GS4, and the group of canonical dimers, *e.g.* lentil lectin). The monomeric Arc5 struc-

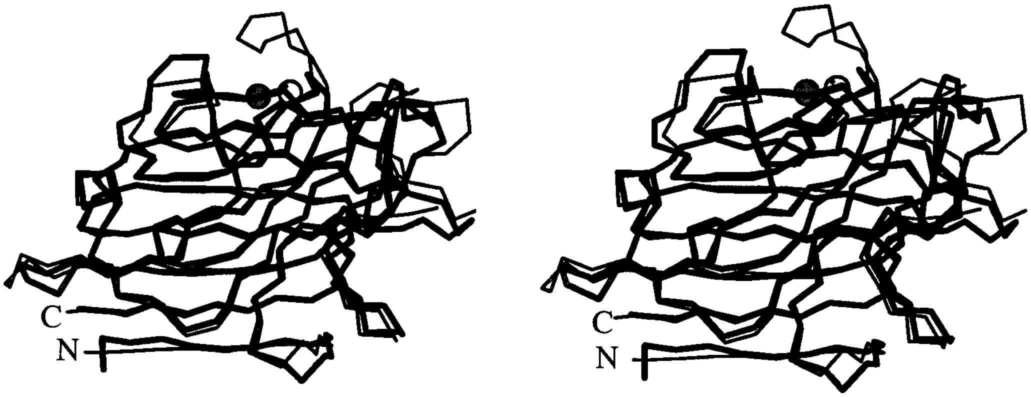

FIG. 2. **The superposition of Arc5 on the PHA-L monomer.** The Arc5 backbone is traced with a *thick line*, and the PHA-L backbone is traced with a *thin line*. The two bound metals of PHA-L, Ca^{2+} and Mn^{2+}, are represented as a *gray* and a *white sphere*, respectively. The major difference between the legume lectin monomers and Arc5, *i.e.* the absence of the metal binding loop in Arc5 (*top* of the *figure*), is clearly visible.

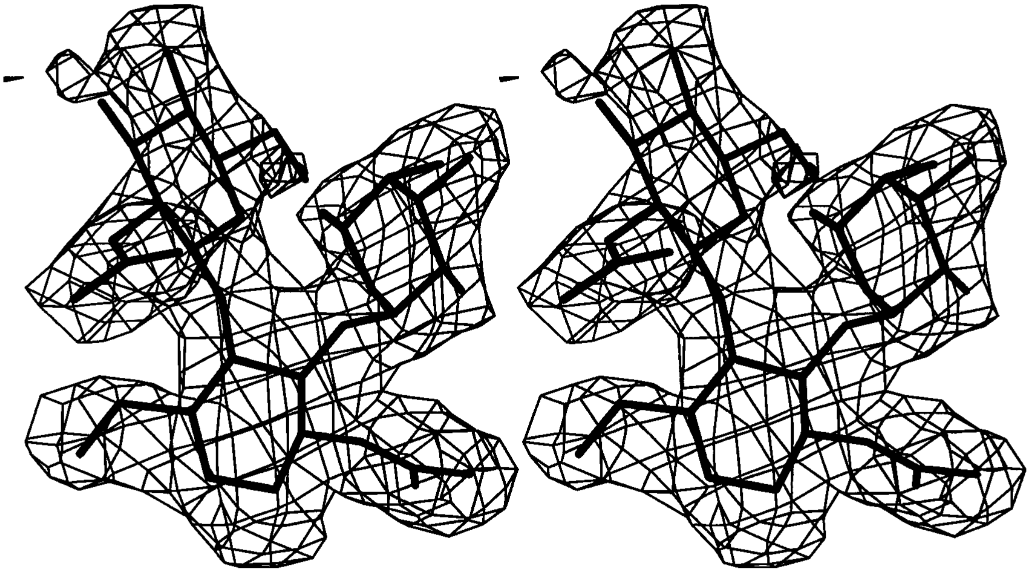

FIG. 3. **The $|F_o - F_c|$ electron density around the glycan bound to Asn^{22} (GlcNAcβ1→4(Fucα1→3)GlcNAc).** The electron density was visualized at the 0.8 σ level.

ture brings the number of known quaternary structures to seven.

The presence of glycans has been invoked to explain the different quaternary structures of GS4 (6) and EcorL (7). 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 (34). 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 of 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 (16) 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 residue in PHA-L.

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 Arc5a contains two glycans, Arc5b contains one, and Arc5c contains none (15).

Interpretable density was only observed for the glycan attached to Asn^{22} (see Fig. 3). Asn^{22} is located in the loop that connects the first β-strand of the front sheet and the second 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 Arc5a 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 sufficiently visible. The electron density of these three visible sugar residues (GlcNAcβ1→4(Fucα1→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 (6) and EcorL (7) 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 (Manα1→6(Manα1→3)(Xylβ1→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 GlcNAcβ1→4(Fucα1→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.

Table III
The φ and Ψ torsion angles (φ/Ψ) of the GlcNAcβ1→4(Fucα1→3) GlcNAc core of the glycan moieties of GS4, EcorL, Arc5, and bromelain (the latter) determined by NMR measurements on the glycan in solution

	GS4	EcorL	Arc5	Bromelain
		degrees		
Fucα1→3GlcNAc	30/18	45/19	48/10	45/30
GlcNAcβ1→4GlcNAc	51/7	47/11	34/3	50/10

Bouwstra *et al.* (24) 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 glycosidic bonds of the three-residue core, which the glycans from Arc5, GS4, EcorL, and bromelain have in common, are shown in Table III. The ψ and φ torsion angles are defined as (C1-O-Cx-Hx) and (H1-C1-O-Cx), respectively, where x = 3 for Fucα1→3GlcNAc and x = 4 for GlcNAcβ1→4GlcNAc. The difference between the Arc5 glycan and the other glycans lies mainly in the conformation of the Fucα1→3GlcNAc glycosidic bond.

The fact that only the core GlcNAcβ1→4(Fucα1→3)GlcNAc moiety is observed in both Arc5 and GS4 may be due to the inherent rigidity of this moiety. Imberty *et al.* (25, 26) studied the conformations of the different disaccharides present in glycans using molecular modeling and energy calculations. For each disaccharide, a conformational energy map in function of

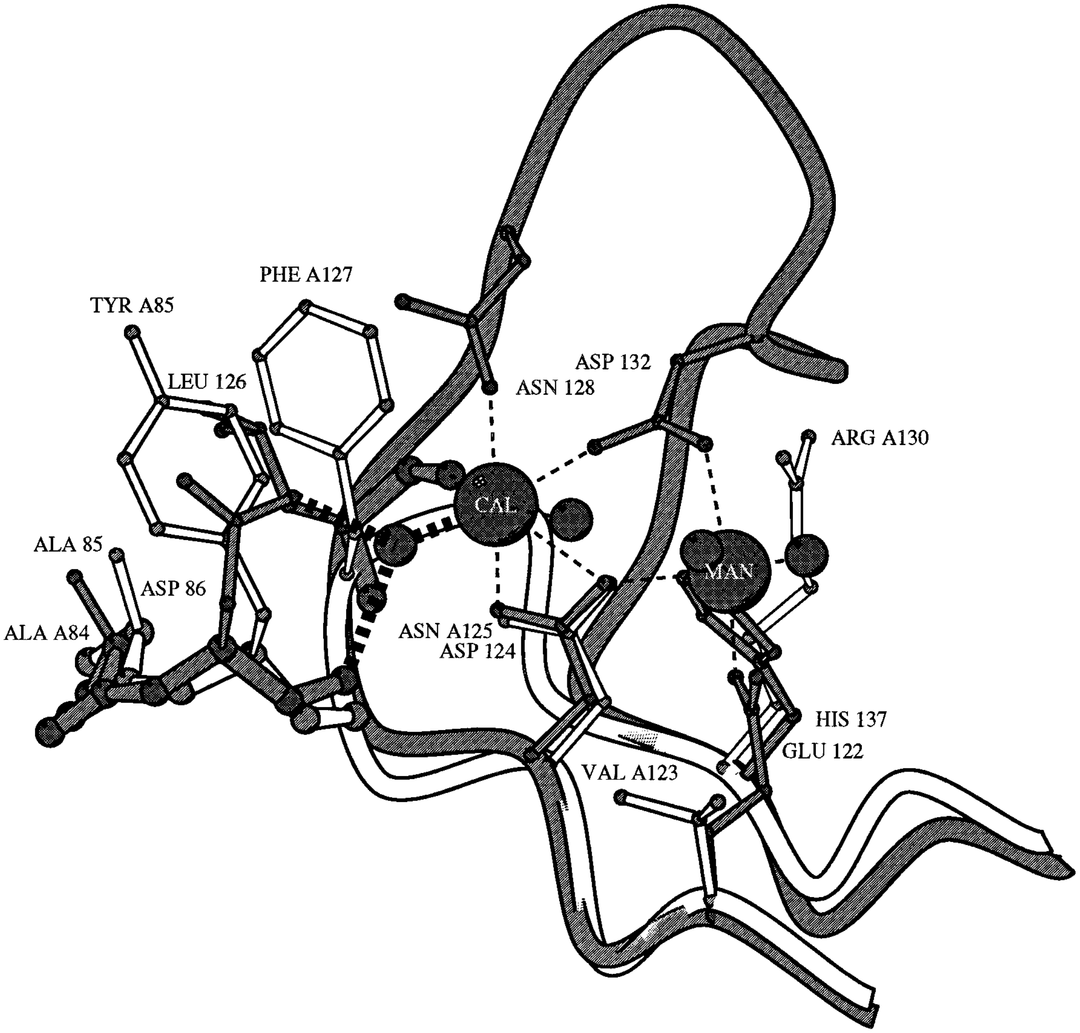

Fig. 4. **The metal binding site of PHA-L superimposed on the truncated metal binding site of Arc5.** The backbone of Arc5 is outlined in *white*, the backbone of PHA-L in *gray*. The amino acids belonging to Arc5 are labeled with an *A*. Only the side chains of the relevant amino acids involved in PHA-L 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 with *thick dotted lines*. This interaction stabilizes the *cis*-peptide bond between Ala^{85} and Asp^{86} in PHA-L. The *figure* clearly shows the stacking interaction between Tyr^{85} and the Phe^{127} residue in Arc5.

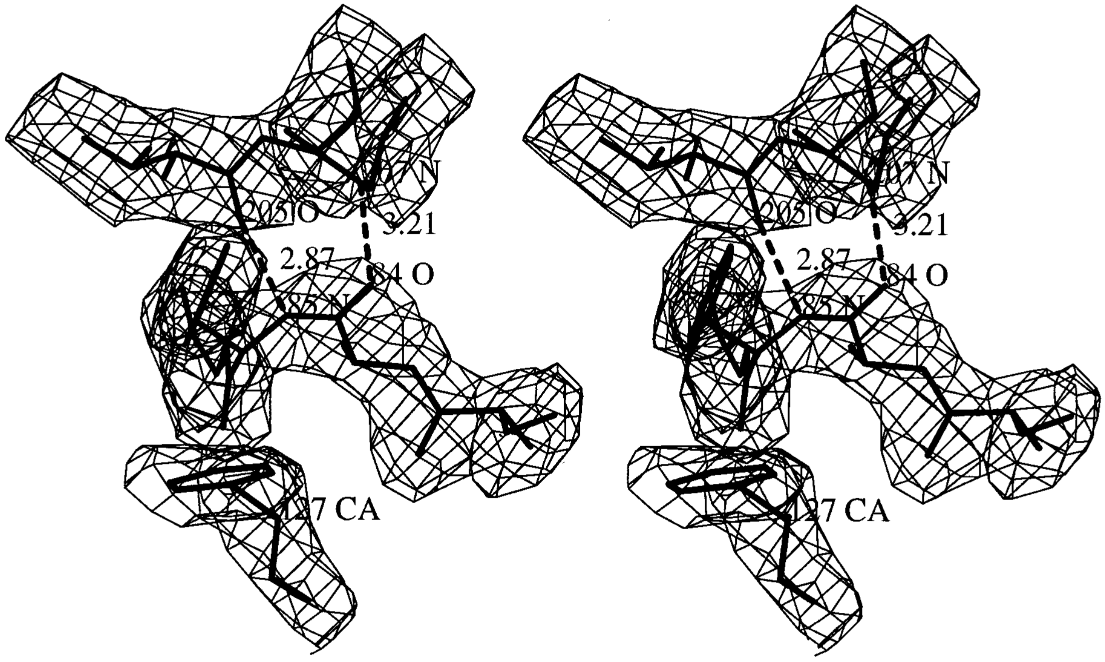

FIG. 5. **Stereo figure of the $|F_o - F_c|$ SA omit map around the two residues involved in the *cis*-peptide bond, Ala^{84} and Tyr^{85}, and the three residues that stabilize the *cis*-peptide bond through stacking interactions with the Tyr side chain (Phe^{127}) or hydrogen bonds (Thr^{205} and Gly^{207}).** The map is visualized at the 1.5 σ level and clearly shows the presence of the *cis* conformation.

the φ and ψ values was calculated. For the GlcNAcβ1→4GlcNAc disaccharide, six low energy conformations were found (A1–A6) (25). 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 Fucα1→3GlcNAc disaccharide, only three energy minima were found L1–L3 (26). In general, this disaccharide is conformationally more restricted then the GlcNAcβ1→4GlcNAc disaccharide. Again the φ/ψ pairs of the four glycans cluster around a reported energy minimum (L1).

It can be concluded that the results about the GlcNAcβ1→4-(Fucα1→3)GlcNAc conformation obtained from NMR measurements in solution, molecular modeling, and x-ray crystallography are in close agreement with each other.

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. 4). 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).

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 (9). 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^{85} and Tyr^{85}, despite the absence of the two bound metal ions. The *cis*-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 Tyr^{85} 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 often involved in hydrophobic interactions with the bound sugar.

Fig. 5 shows an $|F_o - F_c|$ 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.* (28), only 0.36% of all of the peptide bonds in the set of proteins studied were 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 (28) 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% were found to be *cis*. These rare non-*X*-Pro *cis*-peptide bonds are often involved in catalysis or play some other important functional role (29). As far as we know, this is the first report of an Ala-Tyr *cis*-peptide bond.

At present, the mechanism behind the toxicity of the arcelins toward 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 *P. vulgaris* and a fifth arcelin-like protein from *Phaseolus acutifolius* (27)) the Ala residue followed by an aromatic residue (Phe in the case of Arc4; Tyr in the four other cases) is found.

Concanavalin B (ConB), 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 ConB 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 (30). Yet ConB 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 ConB (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 ConB protein (30).

CONCLUSIONS

The structure of arcelin-5, a defense protein from *P. vulgaris* related to the legume lectin class, has been solved at a resolution of 2.7 Å. Arc5 crystallized as a monomer whose structure is similar to the structure of the legume lectin monomers, although the metal binding loop is deleted. One of the most striking features of this structure is the conservation of a *cis*-peptide bond that is found in all of the legume lectins, despite the absence of a major portion of the metal binding site that stabilizes this bond in the legume lectins. While this *cis*-peptide bond is always between an Ala and an Asp residue in the 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* (28) and that these *cis*-peptide bonds often fulfill some functional role (29), the conservation of this feature suggests that it could be involved in the (unknown) mode of action of the protein. We hope that the availability of this structure will stimulate further research on the precise mechanism behind the toxicity toward insects of this interesting protein.

Acknowledgments—We thank Dr. J. Bouckaert and Dr. J. Doumen.

REFERENCES

1. Sharon, N., and Lis, H. (1990) *FASEB J.* **4,** 3198–3208
2. Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E., and Suddath, F. L. (1986) *J. Biol. Chem.* **261,** 16518–16527
3. Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R., and Wyns, L. (1993) *Biochemistry* **32,** 8772–8781
4. Bourne, Y., Mazurier, J., Legrand, D., Rougé, P., Montreuil, J., Spik, G., and Cambillau, C. (1990) *Proteins Struct. Funct. Genet.* **8,** 365–376
5. Bourne, Y., Mazurier, J., Legrand, D., Rougé, P., Montreuil, J., Spik, G., and Cambillau, C. (1994) *Structure* **2,** 209–219
6. 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., and Lemieux, R. U. (1989) *Trans. ACA* **25,** 65–76
7. Shaanan, B., Lis, H., and Sharon, N. (1991) *Science* **254,** 862–866
8. Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A., and Edelman, G. M. (1975) *J. Biol. Chem.* **250,** 1513–1524
9. 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. U. S. A.* **91,** 227–231
10. Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F., and Sacchettini, J. C. (1995) *Biochemistry* **34,** 4933–4942
11. Hamelryck, T. W., Dao-Thi, M., Poortmans, F., Chrispeels, M. J., Wyns, L., and Loris, R. (1996) *J. Biol. Chem.* **271,** 20479–20485
12. Chrispeels, M. J., and Raikhel, N. V. (1991) *Plant Cell* **3,** 1–9
13. Diaz, C., Melchers, L., Hooykaas, P. J. J., Lugtenberg, B. J. J., and Kijne, J. W. (1989) *Nature* **338,** 579–581
14. Kleywegt, G. J., and Jones, T. A. (1995) *Structure* **3,** 535–540
15. Goossens, A., Geremia, R., Bauw, G., Van Montagu, M., and Angenon, G. (1994) *Eur. J. Biochem.* **225,** 787–795
16. Osborn, T. C., Alexander, D. C., Sun, S. S. M., Cardona, C., and Bliss, F. A. (1988) *Science* **240,** 207–210
17. Pflugrath, J. W., and Messerschmidt, A. (1989) *MADNESS Manual of FAST Diffractometer*, Enraff Nonius, Delft, The Netherlands
18. Collaborative Computing Project 4, SERC Daresbury Laboratory (1994) *Acta Crystallogr. Sec. D* **50,** 760–763
19. Navaza, J. (1994) *Acta Crystallogr. Sec. A* **50,** 157–163
20. Peitsch, M. C. (1996) *Biochem. Soc. Trans.* **24,** 274–279
21. Brünger, A. T. (1992) *X-PLOR, version 3.1: A System for Crystallography and NMR*, Yale University, New Haven, CT
22. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
23. 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
24. Bouwstra, J. B., Spoelstra, E. C., De Waard, P., Leeflang, B., Kamerling, J. P., and Vliegenthart, J. F. G. (1990) *Eur. J. Biochem.* **190,** 113–122
25. Imberty, A., Gerber, S., Tran, V., and Pérez, S. (1990) *Glycoconjugate J.* **7,** 27–54
26. Imberty, A., Delage, M., Bourne, Y., Cambillau, C., and Pérez, S. (1991) *Glycoconjugate J.* **8,** 456–483
27. Mirkov, T. E., Wahlstrom, J. M., Hagiwara, K., Finardi-Filho, F., Kjemstrup, S., and Chrispeels, M. J. (1994) *Plant Mol. Biol.* **26,** 1103–1113
28. Stewart, D. E., Sarkar, A., and Wampler, J. E. (1990) *J. Mol. Biol.* **214,** 253–260
29. Herzberg, O., and Moult, J. (1991) *Proteins Struct. Funct. Genet.* **11,** 223–229
30. Hennig, M., Jansonius, J. N., Terwisscha van Scheltinga, A. C., Dijkstra, B. W., and Schlesier, B. (1995) *J. Mol. Biol.* **254,** 237–246
31. Young, M. N., and Oomen, R. P. (1992) *J. Mol. Biol.* **228,** 924–934
32. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
33. Kornegay, J., Cardona, C., and Posso, C. (1993) *Crop Sci.* **33,** 589–594
34. Hartweck, L., Vogelzang, R., and Osborn, T. (1991) *Plant Physiol.* **97,** 204–211
35. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sec. A* **47,** 110–119