Crystal Structure of the Arcelin-1 Dimer from Phaseolus vulgaris at 1.9-Å Resolution

Lionel Mourey, Jean-Denis Pedelacq, Catherine Birck, Christine Fabre, Pierre Rougé, Jean-Pierre Samama

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 273, No. 21, Issue of May 22, pp. 12914–12922, 1998

Printed in U.S.A.

Crystal Structure of the Arcelin-1 Dimer from *Phaseolus vulgaris* at 1.9-Å Resolution*

(Received for publication, December 5, 1997, and in revised form, February 25, 1998)

Lionel Mourey‡, Jean-Denis Pédelacq‡, Catherine Birck‡, Christine Fabre§, Pierre Rougé§, and Jean-Pierre Samama‡¶

From the ‡Groupe de Cristallographie Biologique, and the §Groupe Lectines et Reconnaissance, Institut de Pharmacologie et de Biologie Structurale, UPR 9062 CNRS, 205 route de Narbonne, F-31077 Toulouse CEDEX, France

Arcelin-1 is a glycoprotein from kidney beans (*Phaseolus vulgaris*) which displays insecticidal properties and protects the seeds from predation by larvae of various bruchids. This lectin-like protein is devoid of monosaccharide binding properties and belongs to the phytohemagglutinin protein family. The x-ray structure determination at 1.9-Å resolution of native arcelin-1 dimers, which correspond to the functional state of the protein in solution, was solved using multiple isomorphous replacement and refined to a crystallographic *R* factor of 0.208. The three glycosylation sites on each monomer are all covalently modified. One of these oligosaccharide chains provides interactions with protein atoms at the dimer interface, and another one may act by preventing the formation of higher oligomeric species in the arcelin variants. The dimeric structure and the severe alteration of the monosaccharide binding site in arcelin-1 correlate with the hemagglutinating properties of the protein, which are unaffected by simple sugars and sugar derivatives. Sequence analysis and structure comparisons of arcelin-1 with the other insecticidal proteins from kidney beans, arcelin-5, and α-amylase inhibitor and with legume lectins, yield insights into the molecular basis of the different biological functions of these proteins.

Insecticides are widely used to protect crop plants against their natural predators. Nevertheless, the extensive use of these compounds has given rise to widespread concern about insect resistance and soil pollution. This has prompted active investigation into new protective strategies, and the development of genetically engineered crops expressing insecticidal proteins appears attractive (1). The seeds of the kidney bean (*Phaseolus vulgaris*) contain proteins encoded by four tightly linked genes (2), generally referred to as the phytohemagglutinin (PHA)[1] family of bean proteins. Two of these polypeptides (E and L) lead to all possible combinations of the tetrameric assembly of the PHA lectins (E_4, E_3L, E_2L_2, L_3E, and L_4). The products of the two other genes, α-amylase inhibitor (α-AI) and arcelin, which were named lectin-like proteins, display insecticidal properties (3, 4).

The PHA family belongs to the superfamily of legume lectins, in which the protein subunits are equivalent in size, display significant sequence homology, and have a common β-sandwich fold. However, these closely related proteins differ in their glycosylation patterns, quaternary structure organization, and sugar-binding specificities (Table I). Several types of dimers and tetramers have been characterized in the legume lectins, and there seems to be no apparent relationship between sugar-binding properties, oligomerization, and glycosylation states. Mono- or disaccharides bind to legume lectins at a well defined site on each subunit. Sugar recognition involves common structural and sequence environments. They include a conserved core that provides binding energy irrespective of specificity, a peptide shell around this core which defines monosaccharide specificity, and an outer hypervariable region that may be responsible for interaction with oligosaccharides (20–21). Based on sequence alignment, α-AI and arcelins contain substitutions and/or deletions of essential amino acid residues involved in this molecular recognition (Fig. 1), which likely explains why these proteins do not bind simple sugars (19, 22). However, the hemagglutinating properties toward protease-treated erythrocytes (19, 23) and the specific interactions of dimeric arcelin-1 with glycoproteins (19) is consistent with the presence of a complex oligosaccharide-binding site on the protein. Contrastingly, no lectin activity has been reported for α-AI.

α-AI inhibits α-amylases of mammalian and insect origin, but has no effect on the plant enzymes (24). The expression of this protein in tobacco (25) and pea (3, 26) caused these transgenic plants to become resistant against some insect pests. Arcelin exists as six electrophoretic variants, and the most promising ones conferring insect resistance are arcelin-1 and arcelin-5 (27). The insecticidal properties of these glycoproteins, which are lethal to the larvae of bruchids, appear different to that of α-AI because arcelin displays no inhibitory properties toward α-amylase (19). These larvae invade and damage the seeds of economically important crops such as soybean (*Glycine max*), cowpea (*Vigna unguiculata*), kidney bean, pea (*Pisum sativum*), and lentil (*Lens culinaris*). This suggests that the seeds of transgenic crops harboring these proteins might be protected from attack by insect pests. One of the requirements to be fulfilled for any broad development of these genetically engineered plants is a detailed analysis of the functional properties of arcelins.

* This work was supported in part by the the Conseil Régional de Midi-Pyrénées (contract number RECH 9609713). 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 (codes 1AVB and R1AVBSF) have been deposited in the Protein Data Bank, Brookhaven National Laboratory, Upton, NY.

¶ To whom correspondence should be addressed: IPBS-CNRS, 205 route de Narbonne, 31077 Toulouse Cedex, France. Tel.: 33 5 61 17 54 44; Fax: 33 5 61 17 54 48; E-mail: samama@ipbs.fr.

[1] The abbreviations used are: PHA, phytohemagglutinin; α-AI, α-amylase inhibitor; ConA, *Canavalia ensiformis* concanavalin A; EcorL, *Erythrina corallodendron* lectin; Gal, galactose; Glc, glucose; GS4, *Griffonia simplicifolia* isolectin IV; LoLI, *Lathyrus ochrus* isolectin I; Man, mannose; GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; NCS, noncrystallographic symmetry; MIR, multiple isomorphous replacement; r.m.s., root mean square.

TABLE I
Legume lectins and lectin-like glycoproteins of known three-dimensional structure

	Abbreviation	Glycoprotein	Type of specificity	Quaternary structure	Methods[a]	References to x-ray structure
Lectins						
Concanavalin A	ConA	no	Man/Glc	222 tetramer	MIR	(5)
Pea lectin	PsA	no	Man/Glc	Canonical dimer I	MIR	(6)
Favin		yes	Man/Glc	Canonical dimer I	MR	(7)
Lathyrus ochrus isolectin I	LoLI	no	Man/Glc	Canonical dimer I	MR	(8)
Erythrina corallodendron lectin	EcorL	yes	Gal	Dimer II[b]	MR	(9)
Griffonia simplicifolia isolectin IV	GS4	yes	Gal, Lewis b and Y blood determinants	Dimer III[b]	MR	(10)
Lentil lectin	LcA	no	Man/Glc	Canonical dimer I	MR	(11)
Peanut agglutinin	PNA	no	Gal-GalNAc	"Open" tetramer	MIR	(12)
Soybean agglutinin	SBA	yes	GalNAc/Gal	222 tetramer	MR	(13)
Phytohemagglutinin-L	PHA-L	yes	Pentasaccharide	222 tetramer	MR	(14)
Vicia villosa isolectin B4	VVLB4	yes	GalNAc/Gal, Tn glycopeptide	222 tetramer	MR	(15)
Lectin-like						
α-amylase inhibitor (in complex with α-amylase)	α-AI	yes	No reported lectin activity	Canonical dimer I	MR	(16)
Arcelin-5	Arc5	yes	No or very low lectin activity (17)	Monomer	MR	(18)
Arcelin-1	Arc1	yes	Complex glycans (19)	Canonical dimer I[b]	MIR	(this work)

[a] MIR, multiple isomorphous replacement; MR, molecular replacement.
[b] The three topologies are described under "Results and Discussion."

Solution-state characterization, crystallization of the native arcelin-1 dimer from kidney beans, and preliminary x-ray analysis were described previously (28). In this paper, we report the three-dimensional structure of the protein refined to 1.9 Å resolution and discuss the structural features related to its biochemical properties.

EXPERIMENTAL PROCEDURES

Structure Determination—Crystals of arcelin-1 belong to the orthorhombic space group $P2_12_12$ with cell parameters a = 85.6 Å, b = 92.6 Å, and c = 67.3 Å and diffract to 1.9-Å resolution (28). Two monomers of arcelin-1 were found in the asymmetric unit using molecular replacement and one monomer of LoLI as search model (28). The noncrystallographic symmetry operation, which relates the two monomers, is defined by the direction cosines (0.00098,1.00000,0.00141) of the noncrystallographic axis going through the point of coordinates (46.98, 64.07, 50.46) and a κ angle value of 179.95°. A xenon derivative of arcelin-1 was prepared using the pressurization device and method of Schiltz *et al.* (29). The crystal was equilibrated under a xenon pressure at 15×10^5 pascal for 1 h before data collection. The intensities were collected at 4 °C on a 30-cm Mar imaging plate on beam line DW32 at the LURE synchrotron using a wavelength of 0.975 Å. The anomalous contribution for xenon is small at this wavelength (29), and no attempt was made to collect anomalous data. Diffraction data of a platinum derivative, prepared by soaking crystals in 5 mM K_2PtCl_4 for 21 h, were measured on a Rigaku RAXIS-II imaging plate system equipped with Yale mirror optics (Molecular Structure Corporation) mounted on a Rigaku RU-300 x-ray generator. Data processing was carried out with the MOSFLM package (30). Unless stated, data reduction and all subsequent crystallographic computations were carried out using the programs from the CCP4 suite (31).

The two heavy atom derivatives were analyzed from difference Patterson maps. Refinement of the heavy atom parameters and phase calculations at 3.1-Å resolution were carried out with MLPHARE (CCP4 suite). The initial multiple isomorphous replacement (MIR) phases were improved by solvent flipping (32) and noncrystallographic symmetry (NCS) averaging using the program DM (33). Molecular envelopes for symmetry averaging were constructed from the molecular replacement model (28) using the program MAMA (34).

Crystallographic Refinement—Model building and manual corrections were carried out on a Silicon Graphics Indigo2 Extreme, using Alberta/Caltech TOM, based on FRODO (35). Structure refinement was performed using the program X-PLOR, Version 3.1 (36), including the low resolution data and applying a bulk solvent correction. A randomly selected data set (2130 reflections) was excluded from refinement and used for analysis of the free R factor (37). In each refinement cycle, simulated annealing from 3000 to 300 K, followed by conventional energy minimization, and individual B factors refinement were applied. The initial model, built in the modified MIR map, was refined to 2.5-Å resolution, with strict application of the NCS for the two subunits in the asymmetric unit. The resolution was then extended to 1.9 Å, and after two additional cycles of refinement, the NCS constraints were released. Solvent molecules were added as neutral oxygen atoms when they appeared as positive peaks above 4.0 σ in the $(F_{obs} - F_{calc})\exp(i\alpha_{calc})$ map and displayed acceptable hydrogen-bonding geometry. Hereafter, the simulated annealing step was performed from 500 to 300 K. A bulk solvent model constructed using Babinet's principle (38) and an overall anisotropic B correction, combined to positional refinement, were applied in the last refinement cycles.

Comparison of Arcelin-1 with Arcelin-5 and Lectin Structures—The Protein Data Bank entries 2CNA, 1FAT, 1IOA, 1LEC, 1LTE, and 1LOE and 1LOB of the Brookhaven National Laboratory were used for the comparison of arcelin-1 with ConA (5), *P. vulgaris* PHA-L (14) and arcelin-5 (18), GS4 (10), EcorL (9), and LoLI (8, 39), respectively. The matrices applied to superimpose these protein structures were derived from the least-squares minimization of the positions of the 89 Cα atoms which belong to the two major conserved β-strands present in all of these proteins.

RESULTS AND DISCUSSION

Structure Determination and Refinement—Heavy atom derivatives were readily obtained and the structure was therefore solved using the multiple isomorphous replacement and density modification methods. The coordinates of two xenon atoms bound to the protein were deduced from the Harker sections of the difference Patterson map. Their positions are related by the 2-fold noncrystallographic symmetry. A lower signal was given by the platinum derivative. The statistics of heavy atom derivatives data and phasing are summarized in Table II. The figure of merit, in the resolution range 15–3.1 Å, was 0.32 (0.43 for centric reflections). The electron density, in the initial NCS averaged map, was well defined for 171 residues and 128 side chains, including the Cys^{144}-Cys^{180} disulfide bridge. After refinement to 2.5 Å, applying strict NCS constraints, 221 residues and 207 side chains were assigned, and the R factor dropped from 0.35 (R_{free} = 0.39) to 0.25 (R_{free} = 0.28). Further refinement steps involved successively (i) the extension of resolution to 1.9 Å, (ii) the release of the noncrystallographic symmetry, and (iii) the introduction of the N-glycosylation moieties and of solvent atoms.

The final model contains 226 residues in each monomer, and consists of 3516 non-hydrogen protein atoms, 112 carbohydrate atoms, 2 sulfate groups, and 230 water molecules. Weak electron density was observed in the region 56–62 and for a few solvent-accessible side chains. Alternate side chain conforma-

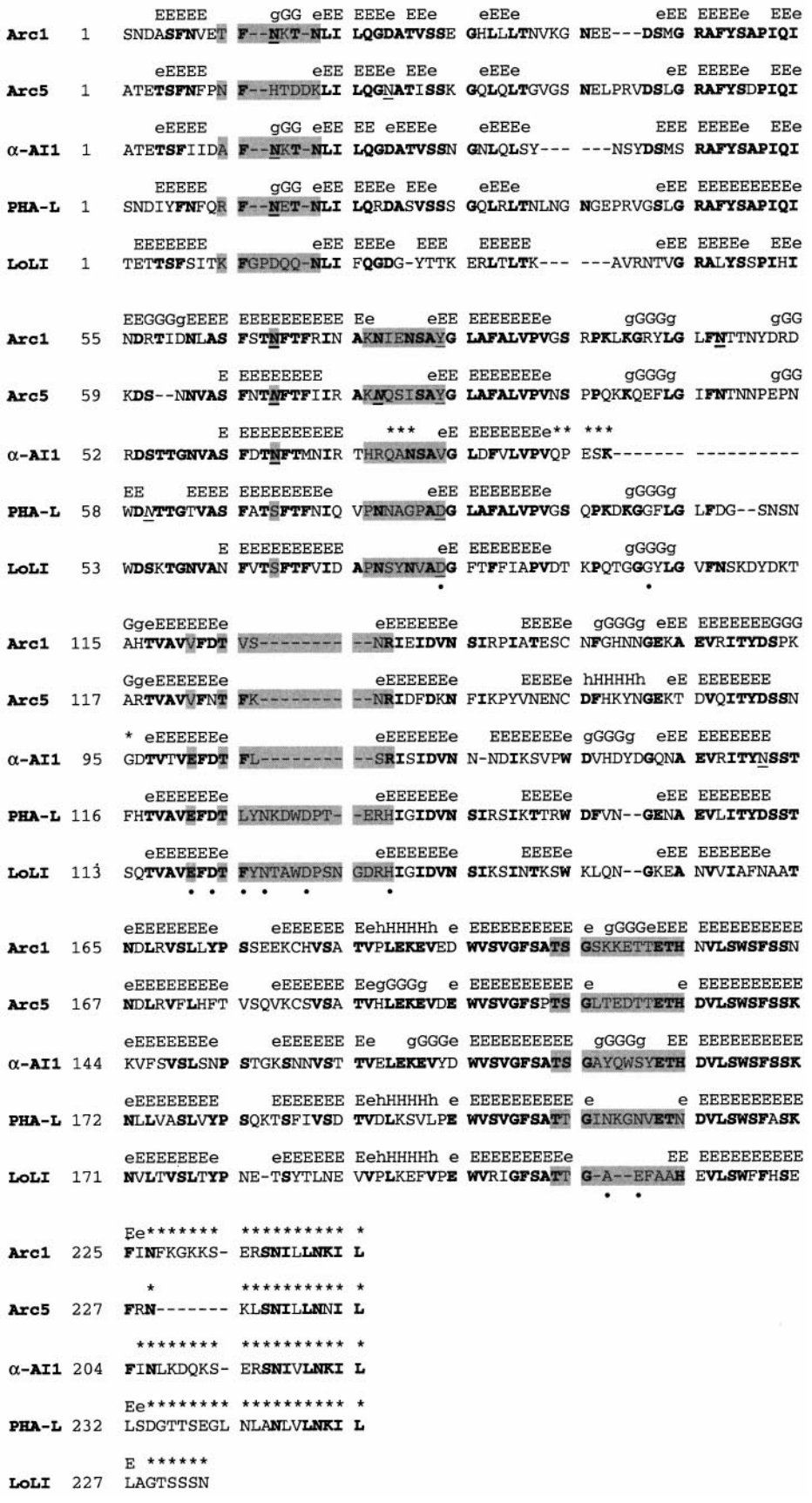

FIG. 1. **Sequence alignment for *P. vulgaris* arcelin-1 (*Arc1*), arcelin-5 (*Arc5*), α-amylase inhibitor 1 (*α-AI1*), phytohemagglutinin-L (*PHA-L*), and *L. ochrus* isolectin I (*LoLI*).** Secondary structures were assessed with the program PROCHECK (40): E for extended strand which participates in β-ladder, G for 3_{10} helix, H for α-helix. e, g, and h denote extension of β-strand, 3_{10} helix, and α-helix, respectively. Other symbols used are: *, corresponds to missing residues in the three-dimensional structures; X (≠N), to *cis*-peptide; N, to glycosylation sites seen in the three-dimensional structures; *N* to putative glycosylation site not seen in the three-dimensional structures. Residues conserved in at least three sequences are shown in *bold*. Amino acid positions involved in metal and monosaccharide binding by LoLI are indicated with a *black dot* below the sequence. Key regions discussed in the text are *shaded*.

tions could be postulated for residue Asn^{110} from both subunits. The final crystallographic *R* factor is 0.208 (R_{free} = 0.242) for 41,590 reflections between 33.71 and 1.9 Å (0.198 and 0.229, respectively, for 35,782 reflections with $F > 3\sigma(F)$). The average *B* factors are 20.9 $Å^2$ for protein atoms (18.0 $Å^2$ and 24.1 $Å^2$ for main chains and side chains, respectively), 38.6 $Å^2$ for sugar atoms, and 32.8 $Å^2$ for solvent atoms. The quality of the stereochemistry was assessed using the program PROCHECK (40).

TABLE II
Statistics of diffraction data and phasing

Data set	Xe	K_2PtCl_4
Resolution range (Å)	31.4–3.1	11.0–3.5
Total observations	27,883	15,511
Unique reflections	8,773	6,260
Completeness (%)	87.5 (87.4)[a]	86.0 (93.3)[a]
R_{merge}[b]	0.102 (0.127)[a]	0.110 (0.126)[a]
Phasing power[c]	1.54/1.20	0.84/0.69
R_{cullis}[d]	0.67 (801)	0.87 (791)

[a] The numbers given in parentheses denote the respective values of the highest resolution shell.

[b] $R_{merge} = \sum_{\mathbf{h}}\sum_{i}\left|\langle I(\mathbf{h})\rangle - I_i(\mathbf{h})\right| \Big/ \sum_{\mathbf{h}}\sum_{i} I_i(\mathbf{h})$, where $I_i(\mathbf{h})$ is the intensity of the ith observation of reflection $\mathbf{h}$ with average intensity $\langle I(\mathbf{h})\rangle$.

[c] Phasing power $= \left[\sum_n |F_H|^2 \Big/ \sum_n |E|^2\right]^{1/2}$, with $\sum_n |E|^2 = \sum_n \left\{\left|F^{PH}_{obs}\right| - \left|F^{PH}_{calc}\right|\right\}^2$ and n is the number of reflections for the derivative. $|F_H|$ is the calculated scattered amplitude of the heavy atom structure; given for acentric and centric reflections, successively.

[d] $R_{Cullis} = \sum_{\mathbf{h}} \left|\left|F_{PH} \pm F_P\right| - F_H\right| \Big/ \sum_{\mathbf{h}} \left|F_{PH} \pm F_P\right|$ for centric reflections (number given in parentheses).

All residues are in the allowed region of a Ramachandran plot (89.3% are in the most favored region). The r.m.s. deviations on bond lengths and bond angles are 0.007 Å and 1.55°, respectively. The upper estimate of the error in the atomic positions from a Luzzati plot (41) lies between 0.15 and 0.25 Å.

Overall Structure—The architecture of the arcelin-1 monomer corresponds to the legume lectin fold and displays the jellyroll Greek key β-barrel motif (Fig. 2). Secondary structure assignment (42) indicated that 130 residues (57.5%) contribute to the formation of two major and of one minor antiparallel β-sheets. The first major sheet consists of six strands: S1 (Ala^4-Val^8), S6 (Asn^{61}-Lys^{76}), S11 (Glu^{152}-Asp^{161}), S12 (Asn^{165}-Tyr^{173}), S13 (Glu^{178}-Val^{186}), and S15 (Thr^{211}-Ile^{226}), and was named sheet I or back sheet in concanavalin A (43). This sheet is flat, and the two longest and adjacent strands (S6 and S15) display a significant curvature at residues 70 and 217, respectively. It is packed against the second major seven-stranded curved sheet: S2 (Asn^{15}-Asp^{21}), S5 (Ser^{42}-Ser^{49}), S7 (Ala^{82}-Val^{92}), S8 (Thr^{117}-Thr^{124}), S9 (Asn^{127}-Asn^{134}), S10 (Ala^{140}-Cys^{144}), and S14 (Asp^{194}-Gly^{205}), called sheet II or front sheet. The minor β-sheet (S3, Thr^{23}-Ser^{25}; and S4, His^{29}-Leu^{32}) is inserted between S2 and S5 of sheet II and stabilizes the curved part of S15 from sheet I. The number of residues in helical conformation is small. They belong to one short α-helix (Pro^{187}-Val^{192}) and to seven 3_{10} helical turns (residues 12–14, 57–60, 99–103, 112–116, 146–150, 162–164, and 207–210). The remaining 63 amino acids (27.9%) are engaged in loops and turns.

The tertiary structure seems stabilized, at one edge of sheets I and II, by the peptide stretch (Ile^{52}-Asp^{56}) that displays hydrogen bonds with strands S6 and S14 and, at the other edge, by a disulfide bridge between residues 144 and 180 from S10 and S13, respectively (Fig. 2). The conformation of the disulfide bridge corresponds to the right-handed spiral with positive χ_2 and χ_3 values (44) and a distance between Cα atoms of 5.7 Å. Legume lectins are known to be poor in sulfur-containing amino acids (45). On the contrary, the cysteine residues at positions 144 and 180 are conserved among arcelin variants 1, 2, and 5 from *P. vulgaris*, and the disulfide bridge reported herein has also been observed in the crystal structure of monomeric arcelin-5 (18). However, only Cys^{144} is conserved in arcelin-4, and the single cysteine residue of arcelin from *Phaseolus acutifolius* is found in a different position.

Only the first 226 amino acids out of the 244 residues deduced from the cDNA sequence of the mature protein (4) have defined electron density. The last amino acid in each of the two arcelin-1 monomers in the asymmetric unit is at the C-terminal end of strand S15, an observation which holds for arcelin-5, α-AI1, and several legume lectins (Fig. 1). This might be due to chain flexibility or to truncation of the C-terminal part of the polypeptide chain, which has been shown to occur during the post-translational modifications of lectins in the ripening seeds (46). In arcelin-1, both events seem possible since (i) differential processing might explain the heterogeneity observed by IEF-PAGE on the protein sample used in this study (19) and (ii) weak density for additional residues consistently appeared for one monomer during refinement.

The Quaternary Structure of Arcelin-1—The two monomers in the asymmetric unit are related by a 2-fold molecular axis. The back β-sheets from both monomers associate to create an extended 12-stranded antiparallel β-sheet that spans the dimer (Fig. 3). Eight hydrogen bonds are exchanged between the main chain atoms from the adjacent S1 strands. Several polar and hydrophobic contacts also contribute to the interface between monomers. These involve the side chains of residues 1–5, 7–10, 12, 14–15, 48–51, and 195, water molecules, and the sugar moieties attached to Asn^{12} (see below). This dimer likely corresponds to the single molecular species found in solution and characterized by biochemical methods and small angle x-ray scattering measurements (19, 28). The change in solvent-accessible surface area (ΔASA) amounts to 2150 $Å^2$ upon dimer formation and involves 10% of the calculated ASA for each monomer. These values are slightly below the average values for homodimers (47–48). However, the arcelin-1 dimer was unaffected by 5 M urea and only partially dissociated by addition of 6 M guanidinium hydrochloride, according to gel filtration analysis (19).

The superposition of the α-carbons of the two subunits gives an r.m.s. difference of 0.11 Å, and there is a strong correlation between the maximum positional deviations and the highest B values along the polypeptide chain, which occur in four loops regions (Fig. 3). Three of these loops (residues 36–41, 76–80, and 206–209) are in the same area, near the monosaccharide-binding site in the homologous lectin structures. The maximum differences (0.45 Å) are found for residues 36–41 and arise from the different crystal packing environments of the two molecules in the asymmetric unit. Loop 76–82 corresponds to the proteolytic processing site of pro-α-AI1 (residues 73–79). Cleavage of this loop at the carboxyl side of Asn^{77} activates α-AI1 as inhibitor of α-amylase (24). The three-dimensional structure of α-AI1 has been solved in complex with pancreatic α-amylase (16), and no electron density could be attributed to residues 75–77 in this protein-protein complex. As already mentioned, arcelin does not inhibit α-amylase. Sequence alignment (Fig. 1) and structure comparison of the two proteins suggests that the type of residue at the N-terminal side of the cleavage site and the occurrence of a *trans*-peptide bond between residues 79 and 80 in α-AI1, which is not found in arcelin-1, might contribute to the specificity of the proteolytic process and thus to the different functions of these proteins.

Despite their high sequence homologies and similar tertiary structures, the proteins of the phytohemagglutinin family display different quaternary structures, which may be a possible factor influencing their biochemical properties. The crystal structure of arcelin-1 represents one example of the so-called "canonical dimer" which has been found for pea lectin (6), favin (7), LoLI (8), lentil lectin (11), and α-AI1 (16). Arcelin-5 from *P. vulgaris*, which may be found as monomers and oligomers in solution (18), displays 62% sequence identity to arcelin-1 and was crystallized as a monomer. The r.m.s. difference between

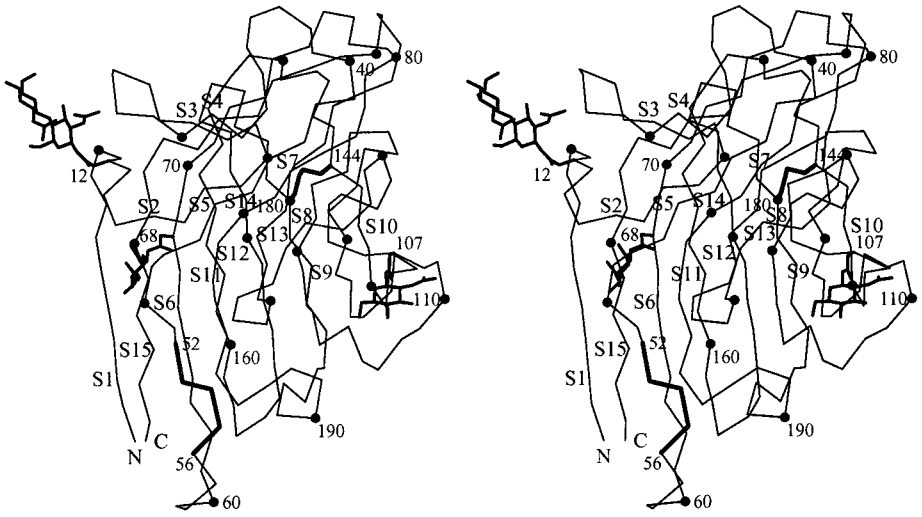

FIG. 2. **Stereoview of the α-carbon trace of one monomer unit of arcelin-1.** Every 10th Cα is indicated by a *black dot*. The *N*-glycosylation sites at Asn[12], Asn[68], and Asn[107], the disulfide bridge between cysteine residues 144 and 180, and residues 52–56 are displayed with *thick lines* and *labeled*. The β-strands are labeled according to the text. Figs. 2–7 were produced using the program MOLSCRIPT (56).

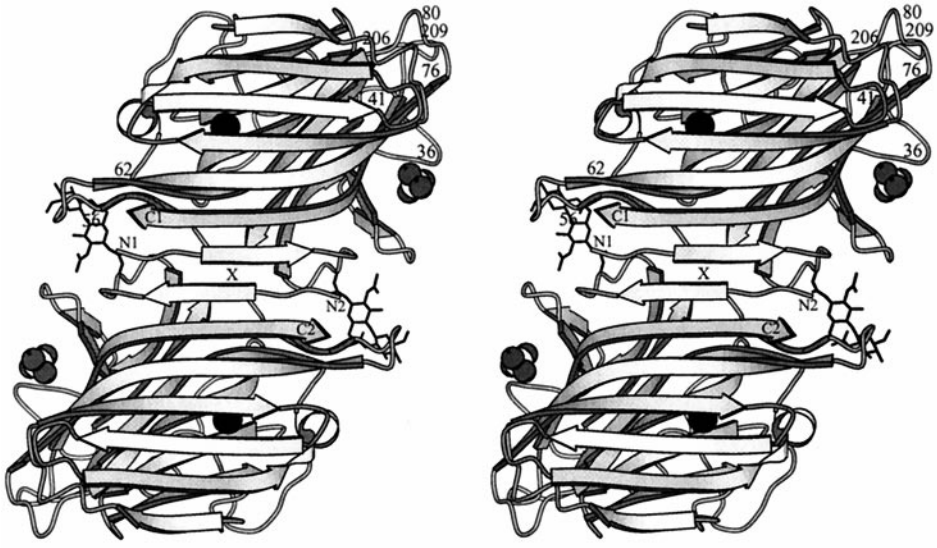

FIG. 3. **Stereodrawing of the native arcelin-1 dimer.** The view is from the 12-stranded dimer-wide β-sheet and down the 2-fold axis (represented by *X*). The xenon-binding sites and the sulfate groups are displayed as CPK spheres (Xe, *black*; O, *gray*; S, *white*). The *N*-glycosylation sites at Asn[12] are displayed with *thick lines*. The N and C termini of both subunits are labeled N1, N2 and C1, C2, respectively. Loops with the highest *B* factors are *numbered* on one monomer.

the positions of the 89 α-carbon atoms in the 13 conserved strands of the two major β-sheets in arcelin-1 and arcelin-5 monomers, is 0.35 Å. Superimposition of these x-ray structures reveals that the different conformation of loop 10–15 in arcelin-5, where one residue is inserted compared with arcelin-1 (Fig. 1), would prevent the formation of an arcelin-1-like dimer. Indeed, severe steric conflicts are observed in this hypothetical dimer of arcelin-5, between Asp[14] and Lys[16] from one monomer, and Trp[197] and Asp[54] from the other monomer, respectively (Fig. 4). The steric conflict arising from Asp[14] might potentially be released by conformational change of the loop 10–14, although accommodation of Lys[16], which belongs to strand S2, cannot be so easily accounted for. This residue is spatially equivalent to Asn[15] in arcelin-1. Here, the side chain of Asn[15] is in close vicinity to that of Val[8] from the same monomer and to the side chain of Asn[2] from the 2-fold related subunit. In arcelin-5, substitution of Val[8] for Phe seems to prevent the Lys[16] side chain from occupying the same position as Asn[15] in arcelin-1. Thus, the formation of arcelin-5 dimers, which were only found to occur in protein fractions which had not been lyophilized (18), should either require conformational rearrangements in this region of the dimer interface, or arise from a different association of the two monomers. The first hypothesis seems more likely since LoLI displays two insertions in the 10–15 region (Fig. 1). Superimposition of LoLI with arcelin-1 and arcelin-5 (r.m.s. difference = 0.45 and 0.48 Å, respectively) shows that the conformation of this loop is different in the three proteins. The second hypothesis seems unlikely based on analysis of the other types of dimers found in EcorL (9) and GS4 (10). Dimerization is achieved through packing of the six-stranded β-sheets (sheet I) from each monomer, with the strands running perpendicularly to each other in GS4, or by building a "handshake"-type interface between the two subunits in EcorL. The sequence of arcelin-5 is 36% identical to those of GS4 and EcorL, and the 89 α-carbon atoms of the conserved strands superimpose with an r.m.s. deviation of 0.55 and 0.58 Å, respectively. Formation of the GS4- and EcorL-type dimers appears to be prevented by substitution of several small hydrophobic or polar residues found at each monomer-monomer interface, by bulky aromatic or charged residues in arcelin-5.

The Glycans Attached to Asn[12], Asn[68], and Asn[107]—Each monomer of arcelin-1 contains 10% carbohydrate (19) and displays three possible *N*-glycosylation sites at Asn[12], Asn[68], and Asn[107], based on the consensus sequence Asn-*X*-Ser/Thr (Fig. 1). The presence of glycan chains at Asn[12] and Asn[107] has been

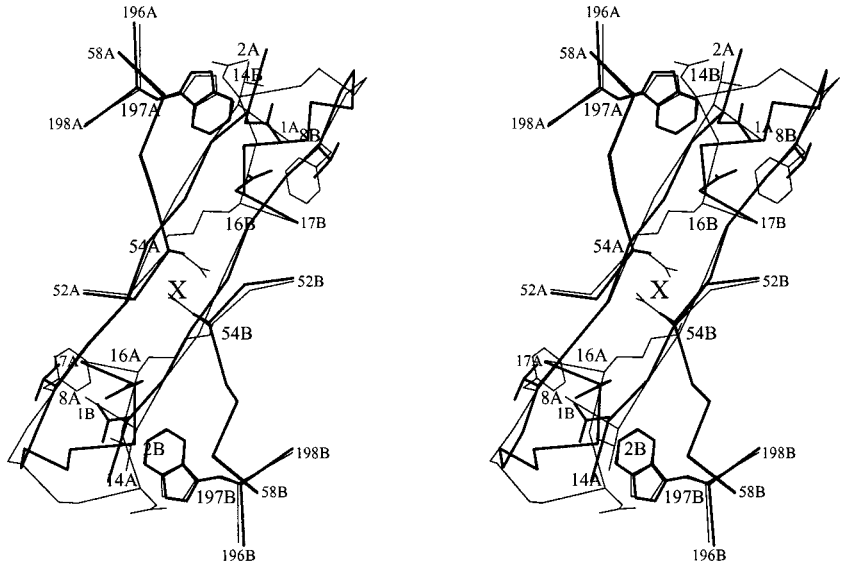

FIG. 4. **Stereoview of the observed arcelin-1 dimer and of the hypothetical arcelin-5 dimer interfaces.** Residues 1–17, 52–58, and 196–198 of arcelin-5 (*thin line*) and residues 1–16, 48–54, and 194–196 of arcelin-1 (*thick line*, two-fold molecular axis represented by X) are displayed. The numbering is shown according to the sequence of arcelin-5, and the two subunits are identified with *A* and *B*, respectively.

demonstrated biochemically (19). The current work, performed on the same protein batch, reveals interpretable electron density for all three glycosylation sites in each subunit. The *N*-linked disaccharide on Asn^{12} is well defined, but only the core GlcNAc residue could be assigned for the two other sites. In all cases, electron density corresponding to additional carbohydrate residues were present but unsuitable for accurate model building. Asn^{12} and Asn^{107} are in solvent-exposed loop regions (Fig. 2). Asn^{68} belongs to strand S6 of sheet I, and the presence of the *N*-acetylglucosamine moiety shows that this location does not prevent the post-translational modification from occurring.

The two β-(1,4)-linked GlcNAc moieties attached to Asn^{12} contribute to the stability of the dimer assembly (Fig. 3) through direct and water-mediated hydrogen bonds to residues 53, 55, 194, and 195 from the 2-fold symmetry-related monomer. The same kinds of interactions were also found in the α-AI1 molecule (16), and the lectin PHA-L is also glycosylated at Asn^{12} (14). Interestingly, these three proteins form "canonical dimers." In PHA-L, the association of two such dimers leads to tetrameric species, although no such oligomeric forms were detected for α-AI1 and arcelin-1 in their solution states. From the current x-ray structure, the formation of tetrameric arcelin-1, with a dimer-dimer interface similar to that found in PHA-L, would be impaired by the glycan chains attached to Asn^{68}. These oligosaccharides face one another in the central channel running between the two dimers and would generate major steric conflicts in a tetrameric assembly. Since α-AI1 and all arcelin variants, but not PHA chains, bear a glycosylation site at a similar location (Fig. 1), it may be that glycosylation of this asparagine could be a factor controlling the formation of higher oligomeric species.

Sulfate-binding Site—The presence of a specific sulfate ion-binding site in arcelin-1 was postulated in order to explain the physicochemical properties in the solution state, and the improved crystallizability of the protein (28). In the refined protein structure, two sulfate ions are bound per dimer. Each binding site is a well defined cleft provided by residues 27–31, 71–74, and 213–217. The anion is bound through a network of polar interactions involving His^{29}, Arg^{72}, and Asn^{215} from one monomer, and Thr^{185} from a crystallographic equivalent of the other monomer (Fig. 5). The requirement of acidic pH for crystallization suggests that His^{29} must be protonated for binding the sulfate ion, which may then promote the dimer-dimer interactions and the growth of single crystals.

Xenon-binding Site—Xenon was shown to provide highly isomorphous derivatives and to bind at the active site of serine proteases (49) and in hydrophobic cavities of proteins (50). Xenon binding in arcelin-1 occurs in a hydrophobic pocket at the interface of the two major β-sheets, at about 10 Å from the protein surface (Fig. 3). The interactions made with protein atoms (Fig. 6) arise from the very high electronic polarizability of the xenon atom which allows attractive van der Waals forces via London interactions (50). The volume of the binding site therefore approximates that of a sphere (about 40 $Å^3$) calculated from the van der Waals radius of xenon (2.16 Å).

The hydrophobic residues that delineate the whole cavity are extremely conserved among ConA and other lectins, and this hydrophobic pocket was involved in the binding of nonpolar molecules, such as iodinated derivatives of aromatic and sugar compounds, and the plant hormone auxin (3-indoleacetic acid) in ConA (51–52). There has been no report of such binding in arcelin-1, and the x-ray structure suggests that the side chains of the hydrophilic residues (Asn^{55}, Arg^{57}, Asp^{60}, Asn^{165}, Glu^{189}) at the entrance of this cavity may prevent the binding of extended hydrophobic probes or of plant hormones.

The Truncated Metal- and Monosaccharide-binding Sites—The structural bases of selective sugar binding by lectins from various origins have been investigated by x-ray structure determinations and were recently reviewed (53). Monosaccharide binding involves four major protein loop segments and two essential Ca^{2+} and Mn^{2+} ions, which bind to the protein approximately 4.5 Å apart, in the conserved core of the lectins. The presence of Mn^{2+} seems important for the proper binding of the Ca^{2+} ion, which in turn makes favorable interactions with a conserved *cis*-peptide bond. Arcelin-1 markedly differs from lectins by the deletion of one of these loops and displays severely impaired monosaccharide binding due to this alteration in the binding site.

Two of the six conserved metal ligands, Asn^{125} and Asp^{129} in LoLI, are in this missing loop of arcelin-1 (Figs. 1 and 7*A*). Two other residues, Glu^{119} and His^{136}, whose side chains are involved in metal binding, are substituted by Val^{121} and Arg^{128} in arcelin-1, respectively. They seem unsuitable for metal coordination, and indeed no bound metal ion was found in arcelin-1 while scrutinizing water molecules and their hydrogen bonding geometries in the course of refinement. The x-ray structures of arcelin-1 and arcelin-5, and sequence alignment considerations suggest that the other arcelin variants should also be devoid of metal ions binding sites in this area.

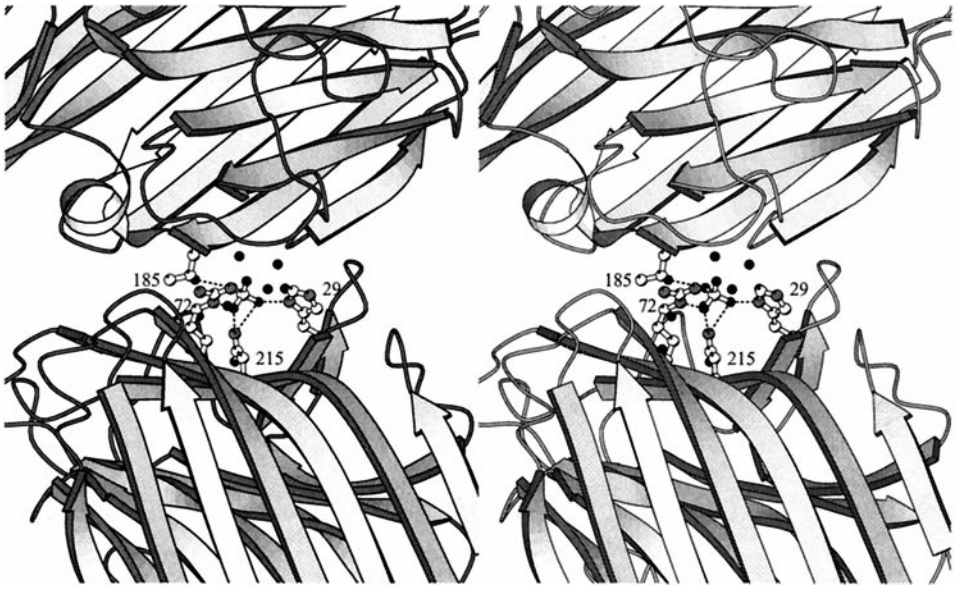

FIG. 5. **Stereoview of the sulfate-binding site of arcelin-1.** Sulfate interactions with protein atoms (≤ 3.1 Å) at the interface between one monomer (*bottom*) and a crystallographic equivalent of the 2-fold symmetry related monomer (*top*) are shown with *dotted lines*. *Gray* and *black* spheres represent nitrogen and oxygen atoms, respectively. Isolated *black spheres* correspond to water molecules.

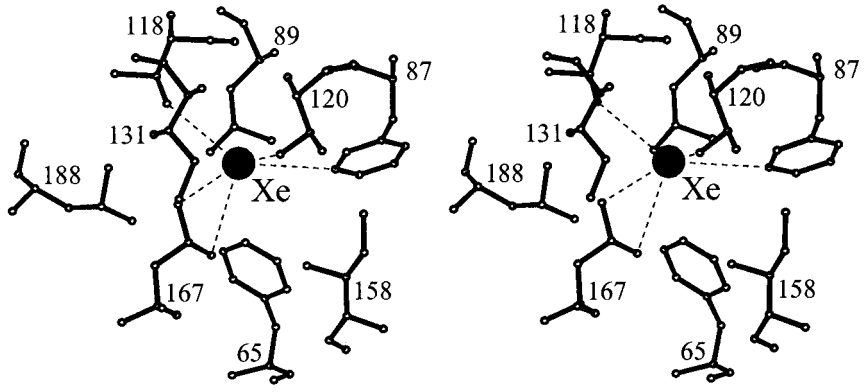

FIG. 6. **Stereoview of the xenon-binding site.** The xenon atom is displayed as a *dark sphere*, and interactions within a 4-Å distance are shown (*dashed lines*).

Arcelin-1 nevertheless displays a *cis*-peptide bond between Ala^{82} and Tyr^{83}, which are spatially equivalent to Ala^{80} and Asp^{81} in LoLI. This conformation, also observed between Ala^{84} and Tyr^{85} in arcelin-5 (18), argues against the proposal that *cis-trans* isomerization of this peptide bond could be induced by metal binding (7, 54). In arcelin-1, the *cis*-peptide bond is stabilized by hydrogen bonds between the main chain nitrogen atom of Tyr^{83} and the main chain oxygen atom of Thr^{203}, and between the main chain oxygen atom of Ala^{84} and the main chain nitrogen atom of Gly^{205}. A third interaction involves the phenolic group of Tyr^{83} and the hydroxyl group of Ser^{206}.

In lectins, the monosaccharide-binding sites form a shallow depression in the vicinity of the cation binding site at the surface of each monomer and display common features for all but one of the four major site-forming loops. The conformation of this loop, which is of variable length, defines the specificity of the sugar binding (10, 55). This topology provides a number of interactions between the protein and the sugar atoms which seem to be impaired in arcelin-1. In the structure of the LoLI-α-methyl-D-mannopyranoside complex (39), the main chain nitrogen atoms of residues 99, 211, and 212 exchange five hydrogen bonds with the sugar. Except for residue 101 of arcelin-1, which lies in approximately the same position as residue 99 in LoLI, the main chain nitrogen atoms of residues 211 and 212 have no counterpart in arcelin-1 due to the different conformation of the region 203–214, where three insertions occur (Figs. 1 and 7). Asn^{125}, which is involved in metal-binding and is hydrogen-bonded to one hydroxyl group of α-methyl-D-mannopyranoside, is deleted in arcelin-1. Finally, Asp^{81} and Phe^{123}, whose side chains provide polar and van der Waals interactions, respectively, with the sugar in LoLI are substituted by Tyr^{83} and Val^{125} in arcelin-1. The tyrosine replaces a residue which is considered to form the basis of the protein-sugar interaction (55), and its side chain, which occupies part of the monosaccharide-binding site, would generate a major steric hindrance against carbohydrate binding to arcelin-1 (Fig. 7*B*).

Conclusions—The three-dimensional structure of arcelin-1 shows that the monomer fold of this lectin-like protein is similar to that of the lectin PHA-L and to the other lectin-like protein from kidney beans, α-AI1. The dimeric structure of arcelin-1 is particularly suited to a function in molecular recognition since it might allow the bridging of cells through interactions with membrane-associated glycoproteins or glycolipids. However, given the rather weak interactions between lectins and monosaccharides (55), the sequence variations and the subsequent structural changes brought to both the metal-binding and combining sites all together explain the loss of monosaccharide-binding activity in arcelin-1 (19). In addition, the steric conflict occurring between a pyranose ring and the side chain of Tyr^{83} should prevent such an interaction from occurring. Along these lines, the weak hemagglutinating property of arcelin-1 toward human and rabbit blood cells was not inhibited by any of the assayed simple sugars and sugar derivatives (19). Nevertheless, the specificity of arcelin-1 for binding

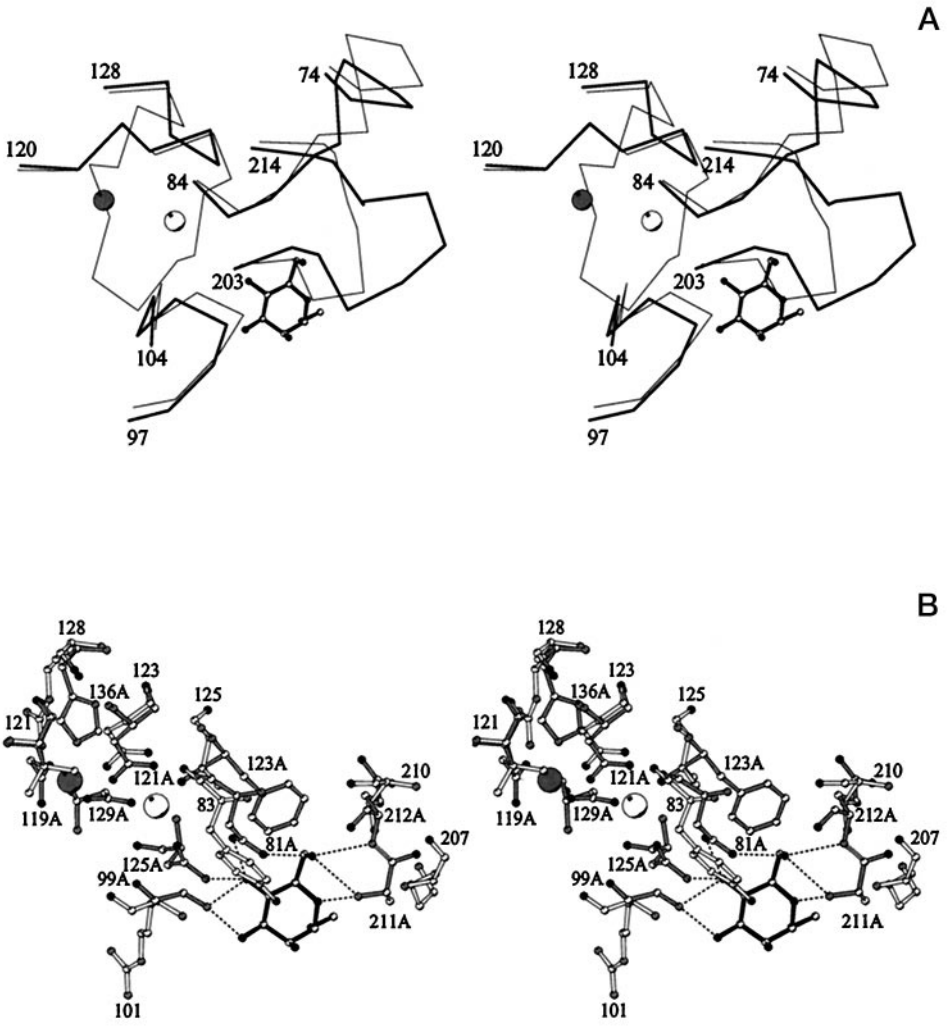

FIG. 7. **Stereoview of the monosaccharide binding site in the LoLI-α-methyl-D-mannopyranoside complex and the corresponding region in arcelin-1 after superimposition of the protein structures.** The carbohydrate is drawn with *black bonds*, and the Ca^{2+} and Mn^{2+} ions are displayed with *large white* and *gray spheres*, respectively. *A*, Cα traces of LoLI (*thin lines*) and of arcelin-1 (*thick lines*). *B*, the side chains involved in metal- and monosaccharide-binding in LoLI (residue numbers with A, *gray bonds*) and the corresponding residues of arcelin-1 (*open bonds*). *White, gray*, and *black* spheres represent carbon, nitrogen, and oxygen atoms, respectively. Hydrogen bond interactions are shown with *dotted lines*.

various glycoproteins, *e.g.* fetuin, asialofetuin, and thyroglobulin, suggests the presence of an extended carbohydrate-binding site in the neighborhood of the unreactive monosaccharide-binding site in arcelin-1, which may recognize the glycan chains of these glycoproteins (19). Structural studies on complexes aimed at mapping this extended sugar-binding site and structure determination of the nonhemagglutinating arcelin-1, which was recently crystallized in our laboratory, should contribute toward an improved understanding of the arcelin-1 function.

Acknowledgments—We thank the scientific staff of LURE (Orsay) for excellent data collection facilities and M. Schiltz and T. Prangé for help with the use of xenon. We are grateful to E. Merritt (University of Washington) for fruitful exchanges about bulk solvent and overall anisotropic *B* correction. We also thank M. Welch for critical reading of the manuscript.

REFERENCES

1. Boulter, D. (1993) *Phytochemistry* **34,** 1453–1466
2. Chrispeels, M. J., and Raikhel, N. V. (1991) *Plant Cell* **3,** 1–9
3. Shade, R. E., Schroeder, 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
4. Osborn, T. C., Alexander, D. C., Sun, S. S. M., Cardona, C., and Bliss, F. A. (1988) *Science* **240,** 207–210
5. Reeke, G. N., Jr., Becker, J. W., and Edelman, G. M. (1975) *J. Biol. Chem.* **250,** 1525–1547
6. Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E., and Suddath, F. L. (1986) *J. Biol. Chem.* **261,** 16518–16527
7. Reeke, G. N., Jr., and Becker, J. W. (1986) *Science* **234,** 1108–1111
8. Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rougé, P., and Fontecilla-Camps, J. C. (1990) *J. Mol. Biol.* **214,** 571–584
9. Shaanan, B., Lis, H., and Sharon, N. (1991) *Science* **254,** 862–866
10. Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Wilson, K. S., and Dauter, Z. (1993) *J. Mol. Biol.* **230,** 950–965
11. Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R., and Wyns, L. (1993) *Biochemistry* **32,** 8772–8781
12. 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
13. Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F., and Sacchettini, J. C. (1995) *Biochemistry* **34,** 4933–4942
14. Hamelryck, T. W., Dao-Thi, M. H., Poortmans, F., Chrispeels, M. J., Wyns, L., and Loris, R. (1996) *J. Biol. Chem.* **271,** 20479–20485
15. Osinaga, E., Tello, D., Batthyany, C., Bianchet, M., Tavares, G., Durán, R., Cerveñansky, C., Camoin, L., Roseto, A., and Alzari, P. M. (1997) *FEBS Lett.* **412,** 190–196
16. Bompard-Gilles, C., Rousseau, P., Rougé, P., and Payan, F. (1996) *Structure* **4,** 1441–1452
17. Goossens, A., Geremia, R., Bauw, G., Van Montagu, M., and Angenon, G. (1994) *Eur. J. Biochem.* **225,** 787–795
18. Hamelryck, T. W., Poortmans, F., Goossens, A., Angenon, G., Van Montagu, M., Wyns, L., and Loris, R. (1996) *J. Biol. Chem.* **271,** 32796–32802
19. Fabre, C., Causse, H., Mourey, L., Koninkx, J., Rivière, M., Hendriks, H., Puzo, G., Samama, J. P., and Rougé, P. (1998) *Biochem. J.* **329,** 551–560
20. Young, N. M., and Oomen, R. P. (1992) *J. Mol. Biol.* **228,** 924–934
21. Adar, R., and Sharon, N. (1996) *Eur. J. Biochem.* **239,** 668–674
22. Rougé, P., Barre, A., Causse, H., Chatelain, C., and Porté, G. (1993) *Biochem. Syst. Ecol.* **21,** 695–703

23. Osborn, T. C., Burow, M., and Bliss, F. A. (1988) *Plant Physiol.* **86,** 399–405
24. Pueyo, J. J., Hunt, D. C., and Chrispeels, M. J. (1993) *Plant Physiol.* **101,** 1341–1348
25. Altabella, T., and Chrispeels, M. J. (1990) *Plant Physiol.* **93,** 805–810
26. Schroeder, H. E., Gollasch, S., Moore, A., Tabe, L. M., Craig, S., Hardie, D. C., Chrispeels, M. J., Spencer, D., and Higgins, T. J. V. (1995) *Plant Physiol.* **107,** 1233–1239
27. Kornegay, J., Cardona, C., and Posso, C. (1993) *Crop Sci.* **33,** 589–594
28. Mourey, L., Pédelacq, J. D., Fabre, C., Causse, H., Rougé, P., and Samama, J. P. (1997) *Proteins* **29,** 433–442
29. Schiltz, M., Prangé, T., and Fourme, R. (1994) *J. Appl. Crystallogr.* **27,** 950–960
30. Leslie, A. G. W. (1992) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography,* **26,** SERC Daresbury Laboratory, Warrington, United Kingdom
31. *Collaborative Computational Project 4* (1994) *Acta Crystallogr. Sec. D* **50,** 760–763
32. Abrahams, J. P., and Leslie, A. G. W. (1996) *Acta Crystallogr. Sec. D.* **52,** 30–42
33. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography,* **31,** SERC Daresbury Laboratory, Warrington, United Kingdom
34. Kleywegt, G. J., and Jones, T. A. (1993) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography,* **28,** SERC Daresbury Laboratory, Warrington, United Kingdom
35. Jones, T. A. (1982) *Computational Crystallography* (Sayre, D., ed) pp. 303–317, Oxford University Press, New York
36. Brünger, A. T. (1992) *X-PLOR,* Version 3.1, Yale University Press, New Haven, CT
37. Brünger, A. T. (1992) *Nature* **355,** 472–475
38. Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G., and Moss, D. S. (1989) *J. Appl. Crystallogr.* **22,** 510–516
39. Bourne, Y., Roussel, A., Frey, M., Rougé, P., Fontecilla-Camps, J. C., and Cambillau, C. (1990) *Proteins* **8,** 365–376
40. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
41. Luzzati, P. V. (1952) *Acta Crystallogr.* **5,** 802–810
42. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22,** 2577–2637
43. Hardman, K. D., and Ainsworth, C. F. (1972) *Biochemistry* **11,** 4910–4919
44. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339
45. Sharon, N., and Lis, H. (1990) *FASEB J.* **4,** 3198–3208
46. 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
47. Jones, S., and Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 13–20
48. Tsai, C.-J., Lin, S. L., Wolfson, H. J., and Nussinov, R. (1997) *Protein Science* **6,** 53–64
49. Schiltz, M., Fourme, R., Broutin, I., and Prangé, T. (1995) *Structure* **3,** 309–316
50. Schiltz, M. (1997) *Utilisation du Xénon et du Krypton pour la Résolution du Probléme des Phases par les Méthods du Remplacement Isomorphe et de la Diffusion Anomale*, Ph.D. thesis, University of Paris XI, Orsay, France
51. Hardman, K. D., and Ainsworth, C. F. (1973) *Biochemistry* **12,** 4442–4448
52. Edelman, G. M., and Wang, J. L. (1978) *J. Biol. Chem.* **253,** 3016–3022
53. Weis, W. I., and Drickamer, K. (1996) *Annu. Rev. Biochem.* **65,** 441–473
54. Brown, R. D., III, Brewer, C. F., and Koenig, S. H. (1977) *Biochemistry* **16,** 3883–3896
55. Rini, J. M. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24,** 551–577
56. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950

Crystal Structure of the Arcelin-1 Dimer from *Phaseolus vulgaris* at 1.9-Å Resolution

Lionel Mourey, Jean-Denis Pédelacq, Catherine Birck, Christine Fabre, Pierre Rougé and Jean-Pierre Samama

J. Biol. Chem. 1998, 273:12914-12922.
doi: 10.1074/jbc.273.21.12914