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

To cite this version:
Lionel Mourey, Jean-Denis Pedelacq, Catherine Birck, Christine Fabre, Pierre Rougé, et al.. Crystal Structure of the Arcelin-1 Dimer from Phaseolus vulgaris at 1.9-Å Resolution. Journal of Biological Chemistry, American Society for Biochemistry and Molecular Biology, 1998, 273 (21), pp.12914-12922. 10.1074/jbc.273.21.12914. hal-03004563

HAL Id: hal-03004563
https://hal-cnrs.archives-ouvertes.fr/hal-03004563
Submitted on 13 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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 Reconnaisance, 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) family of bean proteins. Two of these polypeptides (E and L) lead to all possible combinations of the tetrameric assembly of the PHA lectins (E₄, E₄L, E₃L₂, L₃E, and L₄). 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 superfamilies 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 to 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.
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 P2₁2₁2₁ 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 (59.6, 66.9, 0.464) and a θ angle value of 179.95°. A xenon derivative of arcelin-1 was prepared using the pressurization device and method of Slizt et al. (29). The crystal was equilibrated under a xenon pressure at 15 × 10⁵ pascal for 1 h before data collection. The intensities were collected at 4 °C on a 30-cm Mar imaging plate on beam line D232 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 K₂PtCl₄ for 21 h, were measured on a Rigaku RAXIS-II imaging plate system equipped with 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 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 (Fσ — Fcalc) exp(2πFσ/Fcalc) 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 Arselin-5 and Lectin Structures—The Protein Data Bank entries 2CNA, 1FAT, 1IOA, 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 88 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 Cys⁴⁴⁻Cys⁸⁰ 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 (Rfree = 0.39) to 0.25 (Rfree = 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-
tions could be postulated for residue Asn\textsuperscript{110} from both subunits. The final crystallographic R factor is 0.208 (R\text{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\textsuperscript{3}s(F)). The average B factors are 20.9 Å\textsuperscript{2} for protein atoms (18.0 Å\textsuperscript{2} and 24.1 Å\textsuperscript{2} for main chains and side chains, respectively), 38.6 Å\textsuperscript{2} for sugar atoms, and 32.8 Å\textsuperscript{2} for solvent atoms. The quality of the stereochemistry was assessed using the program PROCHECK (40).

FIG. 1. Sequence alignment for \textit{P. vulgaris} arcellin-1 (\textit{Arc1}), arcellin-5 (\textit{Arc5}), \(\alpha\)-amylase inhibitor 1 (\(\alpha\)-AI1), phytohemagglutinin-L (PHA-L), and \textit{L. ochrus} isoelectin I (LoLI). Secondary structures were assessed with the program PROCHECK (40): E for extended strand which participates in \(\beta\)-ladder, G for 3\text{10} helix, H for \(\alpha\)-helix. e, g, and h denote extension of \(\beta\)-strand, 3\text{10} helix, and \(\alpha\)-helix, respectively. Other symbols used are: *, corresponds to missing residues in the three-dimensional structures; X (\(\neq\)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.
The remaining 63 amino acids (27.9%) are engaged in loops and the minor disulfide bridge corresponds to the right-handed spiral with S10 and S13, respectively (Fig. 2). The conformation of the edge, by a disulfide bridge between residues 144 and 180 from positions 144 and 180 are conserved among arcelin variants 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 Cys144 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-translation 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 Asn12 (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 Å² 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 mM 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 Asn77 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

### Table II

| Data set   | Xe | K₂PtCl₄ |
|------------|----|---------|
| 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) | 86.0 (93.3) |
| Rmerge (%) | 0.102 (0.127) | 0.110 (0.126) |
| Phasing power | 1.54/1.30 | 0.84/0.69 |
| Rmerge (%) | 0.67 (0.81) | 0.57 (0.79) |

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

b Rmerge = \( \sum \frac{\mid I(h) - \langle I(h) \rangle \mid}{\sum I(h)} \)

c Phasing power = \( \sqrt{\frac{\sum \mid F(h) \mid^2}{\sum \mid F(h) \mid^2}} \)

d Rmerge = \( \sum \frac{\mid F(h) - F(h) \mid}{\sum \mid F(h) \mid} \)

and n is the number of reflections for the derivative, F(h) is the calculated scattered amplitude of the heavy atom structure, given for acenetric and centric reflections, successively.

(ground 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 is flat, and the two longest and adjacent strands (S6 and S15) 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 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 Asn12 (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 Å² 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 mM urea and only partially dissociated by addition of 6 M guanidinium hydrochloride, according to gel filtration analysis (19).
the positions of the 89 $\alpha$-carbon atoms in the 13 conserved strands of the two major $\beta$-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 Asp14 and Lys16 from one monomer, and Trp197 and Asp54 from the other monomer, respectively (Fig. 4). The steric conflict arising from Asp14 might potentially be released by conformational change of the loop 10–15, although accommodation of Lys16, which belongs to strand S2, cannot be so easily accounted for. This residue is spatially equivalent to Asn15 in arcelin-1. Indeed, severe steric conflicts are observed in this hypothetical dimer of arcelin-5, between Asp14 and Lys16 from one monomer, and Trp197 and Asp54 from the other monomer, respectively (Fig. 4). The steric conflict arising from Asp14 might potentially be released by conformational change of the loop 10–15, although accommodation of Lys16, which belongs to strand S2, cannot be so easily accounted for. This residue is spatially equivalent to Asn15 in arcelin-1. Here, the side chain of Asn15 is in close vicinity to that of Val18 from the same monomer and to the side chain of Asn15 from the 2-fold related subunit. In arcelin-5, substitution of Val8 for Phe seems to prevent the Lys16 side chain from occupying the same position as Asn15 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 $\beta$-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 $\alpha$-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 Asn12, Asn68, and Asn107—Each monomer of arcelin-1 contains 10% carbohydrate (19) and displays three possible N-glycosylation sites at Asn12, Asn68, and Asn107, based on the consensus sequence Asn-X-Ser/Thr (Fig. 1). The presence of glycan chains at Asn12 and Asn107 has been
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 Asn12 is well defined, but only the core GlcNAc residue could be assigned for the other two sites. In all cases, electron density corresponding to additional carbohydrate residues were present but unsuitable for accurate model building. Asn12 and Asn107 are in solvent-exposed loop regions (Fig. 2). Asn68 belongs to strand S6 of sheet I, and the presence of the N-acetylgalactosamine moiety shows that this location does not prevent the post-translational modification from occurring.

The two β-(1,4)-linked GlcNAc moieties attached to Asn12 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-A1I molecule (16), and the lectin PHA-L is also glycosylated at Asn12 (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-A1I 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 Asn68. 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-A1I 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.

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 Å³) calculated from the van der Waals radius of xenon (2.16 Å).

The hydrophobic residues that delinate 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). Two other residues, Glu119 and His136, whose side chains are involved in metal binding, are substituted by Val121 and Arg128 in arcelin-1, respectively. They seem unsuitable for metal coordination, and indeed no bound metal ion was found in arcelin-1 (18). 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 Ca²⁺ and Mn²⁺ ions, which bind to the protein approximately 4.5 Å apart, in the conserved core of the lectins. The presence of Mn²⁺ seems important for the proper binding of the Ca²⁺ 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.

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 Ca²⁺ and Mn²⁺ ions, which bind to the protein approximately 4.5 Å apart, in the conserved core of the lectins. The presence of Mn²⁺ seems important for the proper binding of the Ca²⁺ 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, Asn125 and Asp129 in LoLI, are in this missing loop of arcelin-1 (Figs. 1 and 7A). Two other residues, Glu119 and His136, whose side chains are involved in metal binding, are substituted by Val121 and Arg128 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.
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-\(\alpha\)-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 \(\alpha\)-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. 7B).

**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, \(\alpha\)-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...
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. Prange 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. Reeye, G. N., Jr., Becker, J. W., and Edelman, G. M. (1975) J. Biol. Chem. 250, 1525–1547
6. Einfahr, H., Parks, E. H., Saguna, K., Subramanian, E., and Suddath, F. L. (1986) J. Biol. Chem. 261, 16158–16157
7. Reeye, G. N., Jr., and Becker, J. W. (1986) Science 234, 1108–1111
8. Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rouge, P., and Fontecilla-Camps, J. C. (1990) J. Mol. Biol. 214, 571–584
9. Shaanan, B., Lis, H., and Sharon, N. (1991) Science 234, 862–866
10. Delhaere, L. T. J., Vandunselwaar, M., Prasad, L., Quail, J. W., Wilson, K. S., and Bouter, 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., Saguna, 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, 4935–4942
14. Hamelryck, T. W., Dao-Thai, M. H., Poortmans, F., Chrispeels, M. J., Wyns, L., and Loris, R. (1996) J. Biol. Chem. 271, 20479–20485
15. Osnaga, E., Tello, D., Battayhany, C., Bianshet, M., Tavares, G., Darin, R., Cervenak, C., Camoin, L., Rosato, A., and Alzari, P. M. (1997) FEBS Lett. 412, 190–196
16. Bompard-Gilles, C., Rousseau, P., Rouge, 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. Young, N. M., and Oomen, R. P. (1992) J. Mol. Biol. 226, 924–934
20. Adar, R., and Sharon, N. (1996) Eur. J. Biochem. 239, 668–674
21. Rouge, P., Barre, A., Causee, H., Chatelain, C., and Porté, G. (1993) Biochim. Biophys. Acta 1195, 336–350
22. Goossens, A., Geremia, R., Bauw, G., Van Montagu, M., and Angenon, G. (1994) Eur. J. Biochem. 225, 787–795
23. Hamelryck, T. W., Poortmans, F., Goossens, A., Angenon, G., Van Montagu, M., Wyns, L., and Loris, R. (1996) J. Biol. Chem. 271, 32796–32802
24. Young, N. M., and Oomen, R. P. (1992) J. Mol. Biol. 226, 924–934
25. Adar, R., and Sharon, N. (1996) Eur. J. Biochem. 239, 668–674
26. Rouge, P., Barre, A., Causee, H., Chatelain, C., and Porté, G. (1993) Biochim. Biophys. Acta 1195, 336–350
27. Goossens, A., Geremia, R., Bauw, G., Van Montagu, M., and Angenon, G. (1994) Eur. J. Biochem. 225, 787–795
28. Hamelryck, T. W., Poortmans, F., Goossens, A., Angenon, G., Van Montagu, M., Wyns, L., and Loris, R. (1996) J. Biol. Chem. 271, 32796–32802
29. Young, N. M., and Oomen, R. P. (1992) J. Mol. Biol. 226, 924–934
30. Adar, R., and Sharon, N. (1996) Eur. J. Biochem. 239, 668–674
31. Rouge, P., Barre, A., Causee, H., Chatelain, C., and Porté, G. (1993) Biochim. Biophys. Acta 1195, 336–350
32. Goossens, A., Geremia, R., Bauw, G., Van Montagu, M., and Angenon, G. (1994) Eur. J. Biochem. 225, 787–795
33. Hamelryck, T. W., Poortmans, F., Goossens, A., Angenon, G., Van Montagu, M., Wyns, L., and Loris, R. (1996) J. Biol. Chem. 271, 32796–32802
34. Young, N. M., and Oomen, R. P. (1992) J. Mol. Biol. 226, 924–934
35. Adar, R., and Sharon, N. (1996) Eur. J. Biochem. 239, 668–674
36. Rouge, P., Barre, A., Causee, H., Chatelain, C., and Porté, G. (1993) Biochim. Biophys. Acta 1195, 336–350
37. Goossens, A., Geremia, R., Bauw, G., Van Montagu, M., and Angenon, G. (1994) Eur. J. Biochem. 225, 787–795
38. Hamelryck, T. W., Poortmans, F., Goossens, A., Angenon, G., Van Montagu, M., Wyns, L., and Loris, R. (1996) J. Biol. Chem. 271, 32796–32802
39. Young, N. M., and Oomen, R. P. (1992) J. Mol. Biol. 226, 924–934
40. Adar, R., and Sharon, N. (1996) Eur. J. Biochem. 239, 668–674
41. Rouge, P., Barre, A., Causee, H., Chatelain, C., and Porté, G. (1993) Biochim. Biophys. Acta 1195, 336–350
42. Goossens, A., Geremia, R., Bauw, G., Van Montagu, M., and Angenon, G. (1994) Eur. J. Biochem. 225, 787–795
43. Hamelryck, T. W., Poortmans, F., Goossens, A., Angenon, G., Van Montagu, M., Wyns, L., and Loris, R. (1996) J. Biol. Chem. 271, 32796–32802
44. Young, N. M., and Oomen, R. P. (1992) J. Mol. Biol. 226, 924–934
45. Adar, R., and Sharon, N. (1996) Eur. J. Biochem. 239, 668–674
23. Osborn, T. C., Burrow, 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, 1239–1239
27. Kornegay, J., Cardona, C., and Posso, C. (1993) Crop Sci. 33, 589–594
28. Mourey, L., Pe`delacq, J. D., Fabre, C., Causse, H., Rou`e, P., and Samama, J. P. (1997) Proteins 29, 433–442
29. Schiltz, M., Prange, 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, 769–783
32. Abrahams, J. P., and Leslie, A. G. W. (1996) Acta Crystallogr. Sec. D. 52, 30–42
33. Kleywegt, G. J., and Jones, T. A. (1993) Acta Crystallogr. Sec. D. 49, 167–339
34. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
35. Bourne, Y., Roussel, A., Frey, M., Rou`e, P., Fontecilla-Camps, J. C., and Camblan, C. (1990) Proteins 8, 356–376
36. Hardman, K. D., and Ainsworth, C. F. (1972) Biochemistry 11, 4910–4919
37. Hardman, K. D., and Ainsworth, C. F. (1973) Biochemistry 12, 4442–4448
38. Edelman, G. M., and Wang, J. L. (1978) J. Biol. Chem. 253, 3016–3022
39. Weis, W. I., and Drickamer, K. (1996) Annu. Rev. Biochem. 65, 441–473
40. Brown, R. D., III, Brewer, C. F., and Koenig, S. H. (1977) Biochemistry 16, 3883–3896
41. Luzzati, P. V. (1952) Acta Crystallogr. 5, 802–810
42. Schiltz, M. (1997) Utilisation du Xe´non et du Krypton pour la Re´solution du Proble`me des Phases par les Me`thodes du Remplacement Isomorphe et de la Diffusion Anomale, Ph.D. thesis, University of Paris XI, Orsay, France
43. 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, 1239–1239
44. Kleywegt, G. J., and Jones, T. A. (1993) Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, 28, SERC Daresbury Laboratory, Warrington, United Kingdom
45. Jones, T. A. (1982) Adv. Protein Chem. 34, 167–339
46. Sharon, N., and Lis, H. (1990) FASEB J. 4, 3198–3208
47. Rini, J. M. (1995) Annu. Rev. Biophys. Biomol. Struct. 24, 551–577
48. 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

Access the most updated version of this article at http://www.jbc.org/content/273/21/12914

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 50 references, 17 of which can be accessed free at http://www.jbc.org/content/273/21/12914.full.html#ref-list-1