Glycobiology, 2021, vol. 31, no. 11, 1543–1556
https://doi.org/10.1093/glycob/cwab059
Advance Access Publication Date: 29 June 2021
Original Article

Structural Biology

Structure and properties of the giant reed (*Arundo donax*) lectin (ADL)

Massimiliano Perduca[1,2], Michele Bovi[2], Laura Destefanis[2], Divina Nadali[2], Laura Fin[2], Francesca Parolini[2], Daniela Sorio[3], Maria E Carrizo[4], and Hugo L Monaco[2]

[2]Biocrystallography and Nanostructure Laboratory, Department of Biotechnology, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy, [3]Centro Piattaforme Tecnologiche, Policlinico G.B. Rossi, Piazzale L.A. Scuro 10, 37134 Verona, Italy, and [4]Centro de Investigaciones en Química Biológica de Córdoba (CIQUIBIC) - CONICET and Departamento de Química Biológica Ranwel Caputto, Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, Haya de la Torre s/n, X5000HUA Córdoba, Argentina

[1]To whom correspondence should be addressed: e-mail: massimiliano.perduca@univr.it

Received 16 September 2019; Revised 15 June 2021; Editorial Decision 15 June 2021; Accepted 15 June 2021

Abstract

Arundo donax lectin (ADL) is a 170 amino acid protein that can be purified from the rhizomes of the giant reed or giant cane by exploiting its selective binding to chitin followed by elution with *N*-acetylglucosamine. The lectin is listed in the UniProt server, the largest protein sequence database, as an uncharacterized protein with chitin-binding domains (A0A0A9P802). This paper reports the purification, structure and ligand-binding properties of ADL. The lectin is a homodimer in which the two protomers are linked by two disulfide bridges. Each polypeptide chain presents four carbohydrate-binding modules that belong to carbohydrate-binding module family 18. A high degree of sequence similarity is observed among the modules present in each protomer. We have determined the X-ray structure of the apo-protein to a resolution of 1.70 Å. The carbohydrate-binding modules, that span a sequence of approximately 40 amino acids, present four internal disulfide bridges, a very short antiparallel central beta sheet and three short alpha helices, two on one side of the beta sheet and one on the other. The structures of the complexes of the lectin with *N*-acetylglucosamine, *N*-acetyllactosamine, *N*-acetylneuraminic acid and *N*-*N*′diacetylchitobiose reveal that ADL has two primary and two secondary carbohydrate-binding sites per dimer. They are located at the interface between the two protomers, and each binding site involves residues of both chains. The lectin presents structural similarity to the wheat germ agglutinin family, in particular, to isoform 3.

Key words: *Arundo donax* lectin, *N*-acetylglucosamine, *N*-acetyllactosamine, *N*-acetylneuraminic acid, *N*,*N*′ diacetylchitobiose

Introduction

Lectins are a protein family of nonimmune origin, which can selectively and reversely bind carbohydrates with a very high degree of specificity and are involved, through sugar-binding, in many basic biological processes (Sharon 2007). Their function is carbohydrate recognition both at the molecular and cellular levels, and they are widespread in all living organisms from viruses to man. Lectins were initially identified in the plant kingdom for their hemagglutinating activity and are now being regularly used in basic and applied research. They are found either in soluble form or membrane-bound and are usually structurally complex molecules with one or more carbohydrate-recognition domains (CRDs) (Taylor and Drickamer 2003).

In plants, lectins have many different important functions, including symbiosis, defense against all sorts of predators like herbivorous insects, and play also a central role in the plant innate immunity defense mechanism against pathogens, such as bacteria, viruses and fungi (De Hoff et al. 2009). Although plant lectins are a relatively

well-studied group, there are still many genes that code for uncharacterized proteins that are probably lectins but still have to be studied and described. Given the level of sophistication of the functions carried out by plant lectins, it is not surprising that the number of lectin families, defined on the basis of structure, sugar specificity and CRDs, is not conclusively established but is in continuous expansion to include always new members (Lannoo and Van Damme 2014).

Several lectins of different origins have been reported to exhibit potent inhibitory effects in vitro on the growth of human tumor cell lines, thus arising much interest in the study of their potential application in the area of diagnosis and therapy of cancer (Gonzàlez De Mejia and Prisecaru 2005). It is also known that lectins can exhibit immunomodulatory effects acting on both the innate and adaptive vertebrate immune systems. These effects depend on the interaction of lectins with glycan moieties present on the surface of immune cells, such as neutrophils, monocytes, macrophages and dendritic cells (DCs), and can lead to enhanced resistance against infection or cancer but also to unwanted effects such as allergy and/or autoimmunity (Souza et al. 2013). The use of nanoparticles coated with lectins that specifically bind to the tumor-associated carbohydrate antigens (TACAs) to target therapeutic drugs to the neoplastic tissues has also been proposed (Devi and Basil-Rose 2018).

Given the importance of lectins in basic and applied research, the identification of novel inexpensive sources and the development of new simple and fast purification protocols is an important area that deserves to be in constant development.

The giant reed or giant cane (*Arundo donax*) is a massive perennial grass native to the Mediterranean basin but very widely diffused to such an extent that it is considered invasive, and its impact on the environment can have very damaging effects on native species and has led to various efforts to reduce its population (Santín-Montanyá et al. 2013). Since it is not known to be a food source for any species or a nesting habitat for wildlife, its best-known practical application seems to be in the energy industry as a bio-fuel crop (Corno et al. 2014). In fact, a study sponsored by the European Union has established *A. donax* as the most productive of all energy biomass crops. During, approximately, 2–3 decades following 1940, it was cultivated in Italy on a large-scale industrial basis as a raw material for the production of paper, but that application appears to be currently no longer economically cost-effective. *Arundo donax* does not normally reproduce through its seeds, which, when present, are usually not fertile, but does it mostly vegetatively by its underground rhizomes, a rich source of starch, proteins and other nutrients, that can penetrate very deeply into the soil and expand very rapidly.

We have used the rhizomes of *A. donax* as the starting material for the purification of a very abundant uncharacterized lectin that can be selectively bound to a chitin affinity column and is eluted with *N*-acetyl-D-glucosamine. The lectin is listed in the UniProt server, the largest protein sequence database, as an uncharacterized protein with chitin-binding domains (A0A0A9P802). A protein which is probably the same that we describe here was purified using an *N*-acetyl-D-glucosamine affinity column and was partially characterized (Kaur et al. 2005).

We report here the characterization of *A. donax* lectin (ADL): purification, cloning, amino acid sequence and 3D structure determination by X-ray diffraction analysis of single crystals of the apo-protein and of its complexes with *N*-acetylglucosamine, *N*-acetyllactosamine, *N*-acetylneuraminic acid and *N*-*N*′ diacetylchitobiose. The lectin exhibits structural similarity to the wheat germ agglutinin (WGA) family, in particular, to isoform 3 (Harata et al. 1995).

Results

Amino acid sequence

The X-ray structure of ADL was determined before its amino acid sequence was known, and therefore, the first information on the sequence was based on the electron density maps of the apo-protein. Given the similarity among the eight modules present in the asymmetric unit, it was relatively straightforward to identify in the UniProt server the protein listed with the label A0A0A9P802, present in the giant cane *A. donax* and cataloged with the name "uncharacterized protein."

The sequence present in the protein database is approximately 85% identical and 93% similar to that of the models we present here, which corresponds to one of the isoforms separated by preparative isoelectric focusing. There are some positions in the chain that show variability, and the UniProt sequence was used to fit the electron density except in those cases where there was discrepancy with the maps.

The amino acid sequences of our model and that of the uncharacterized protein A0A0A9P802 (UNC) are compared in Figure 1. The four modules present in the polypeptide chain have been separated and the last amino acid in each has been labeled in the right-hand column and corresponds to the final sequence in the protein data bank (PDB) file. The amino acids conserved in the four modules are highlighted on a red background and those that are different in the A0A0A9P802 protein are highlighted in green, while the two cysteines that connect the two protomers in the dimer are highlighted in magenta. Using the information present in the UniProt server, the protein was cloned starting from total ribonucleic acid (RNA) extracts of the fresh rhizomes of the plant. The sequence of the complementary deoxyribonucleic acid (DNA) (cDNA) encoding for the protein was then checked and was found to be identical to that deposited.

A final check on the sequence was carried out using mass spectrometry on two different isoforms of ADL following the procedure described below. The results are in agreement with this description and the peaks observed are included in the supplementary materials section (Supplementary Figure S1).

Figure 1 shows also the four disulfide bridges present in each module which were clearly identified in the electron density maps.

Affinity for carbohydrates

The affinity of the lectin for saccharides was studied initially in solution using isothermal titration calorimetry (ITC). Figure 2 shows the results of the ITC experiments with three sugars: *N*-acetylglucosamine, *N*,*N*′ diacetylchitobiose and glucose. Both *N*-acetylglucosamine and *N*,*N*′ diacetylchitobiose display a similar exothermic binding profile with the disaccharide showing a higher affinity (Kd $\sim 4.3 * 10^{-6}$ M). Without imposing any stoichiometry of binding, the standard program used for data analysis of the ITC data finds for the disaccharide two equivalent sites per protein dimer, while glucose does not show any heat of interaction with the lectin. The thermodynamic binding parameters for the two sugars are reported in the supplementary materials section (Supplementary Table SI).

Structure of the protomer

The crystals of ADL belong to space group $P6_5$ with unit cell parameters for the apo-protein $a = b = 127.0$ Å, $c = 46.5$ Å, and they contain one dimer in the asymmetric unit. The stereochemical quality of the protein model was assessed with the program PROCHECK

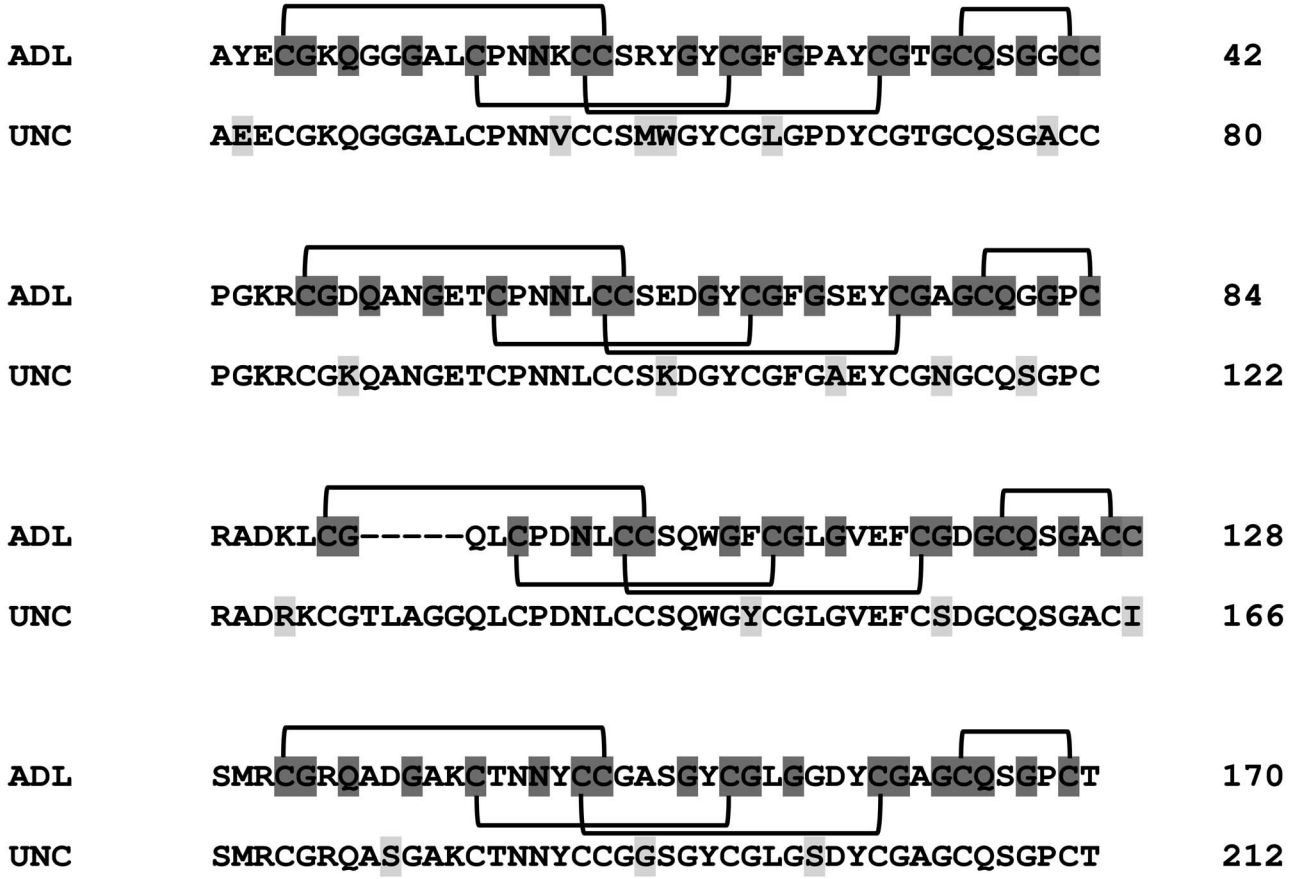

Fig. 1. Amino acid sequence of ADL. Comparison of the amino acid sequence of the models and the sequence deposited in the UniProt server with the labels A0A0A9P802, UNC. The amino acids that are different are highlighted in green and those conserved in the four modules in red. The magenta background identifies the two cysteines that participate in the inter-protomer disulfide bridges in the molecular dimer (Cys 42 and 128). The position in the sequence of the amino acids represented is indicated on the right-hand side of the picture. The four equivalent disulfide bridges present in each domain are also represented. Note the gap after amino acid 92 that corresponds to the disordered region in the electron density maps present in both protomers in the asymmetric unit.

(Laskowski et al. 1993). 86.3% of the residues are in the most favorable region of the Ramachandran plot and the remaining 13.7% in the additionally allowed regions.

The final model of the apo-protein contains 165 amino acids of each protomer, which correspond to a total of 2,250 protein atoms, 1 glycerol molecule and 212 water molecules. There is one area in the map, spanning residues 92–96, where there is no clear interpretable electron density in either of the two protomers, and therefore, that part of the chain was not included in the model. Since the protein that was purified on an *N*-acetyl-D-glucosamine affinity column and partially characterized (Kaur et al. 2005) was reported to be glycosylated, we examined that possibility carefully. Submission of the ADL sequence to a glycosylation prediction server (Chauhan et al. 2013) indicated that there should not be any N-glycosylation sites, but there were two threonines that could potentially be glycosylated. Careful examination of the electron density maps did not indicate that this was indeed the case.

The conventional R factor is 16.0% and the free R factor is 20.6 (Table I). The R factors and root-mean-square deviations (r.m.s.d.) in Table I were calculated with the program Refmac (Murshudov et al. 1997).

Each protomer contains four domains, A-1, A-2, A-3 and A-4 in the first chain and the same number in the identical B chain. They are represented in Figure 3A where they are colored blue, green, orange and red, starting from the N-terminus. The four domains belong to the well-known carbohydrate-binding module family 18 (CBM18), also known as chitin binding 1 module or chitin recognition protein, and are found in a number of lectins that bind *N*-acetylglucosamine (Boraston et al. 2004).

The fold of this domain is also known as the hevein fold (Jiménez-Barbero et al. 2006). In the case of ADL, the domain spans about 40 amino acids organized around a four disulfide bridge core. There are two short central antiparallel beta strands and three short helices, two on one side and one on the other, but the secondary structure is predominantly coil. Figure 3B shows the sequences of the four domains in the protomer aligned as in Figure 1 with the secondary structure elements indicated. The central beta strands are colored in blue and the helices are green. The gap between amino acids 91 and 97 identifies the area in the model that was not built because of the poor quality of the electron density map.

The isolated CBM A-1 is represented in Figure 3C. The N-terminus is identified at the bottom, on the left hand side of the figure, and the cysteines linked by the four yellow disulfide bridges are numbered.

A superposition of the α-carbon models of the four domains in the A protomer colored as in Figure 3A is shown in Figure 3D. The disulfide bridges are not represented for clarity, and the N-terminal end of the peptides is at the bottom left. The amino acids

Table I. Data collection and refinement statistics

Data set	ADL APO	ADL + *N*-acetyl glucosamine	ADL + *N*-acetyllactosamine	ADL + *N*-acetylneuraminic acid	ADL + *N*,*N*′ diacetylchitobiose soaking time 30 s	ADL + *N*,*N*′ diacetylchitobiose soaking time 60 s
Crystal form	1	2	3	4	5	6
Space group	$P6_5$	$P6_5$	$P6_5$	$P6_5$	$P6_5$	$P6_5$
a (Å)	127.04	127.55	126.77	127.21	127.08	126.96
b (Å)	127.04	127.55	126.77	127.21	127.08	126.96
c (Å)	46.49	45.60	46.26	45.37	46.14	45.87
α	90.0	90.0	90.0	90.0	90.0	90.0
β	90.0	90.0	90.0	90.0	90.0	90.0
γ	120.0	120.0	120.0	120.0	120.0	120.0
Resolution range (Å)	27.50–1.70	26.13–1.75	63.38–1.50	63.60–1.60	63.54–1.5	63.48–1.80
Observed reflections	238,287	222,814	411,846	325,854	421,665	239, 931
Independent reflections	47,270	42,913	67,865	55,614	68,370	39,380
Multiplicity*	5.0 (4.8)	5.2 (4.9)	6.1 (5.9)	5.9 (5.9)	6.2 (5.8)	6.1 (5.8)
Rmerge (%)[a]	7.3 (43.5)	9.1 (39.3)	7.3 (44.1)	6.2 (43.4)	6.0 (42.9)	10.8 (56.4)
$<I/\sigma(I)>$	11.3 (3.3)	10.3 (3.4)	10.5 (2.9)	12.0 (3.0)	13.0 (3.1)	10.5 (3.9)
Completeness (%)	99.6 (99.5)	99.8 (99.7)	99.5 (99.6)	99.9 (100.0)	100.0 (99.9)	99.9 (99.8)
Wilson B factor	21.2	19.5	19.0	21.0	20.0	22.8
Reflections in refinement	44,903	40,867	64,365	52,733	64,974	37,303
Rcryst (%)[b]	16.01	14.83	16.23	19.14	15.66	17.41
Rfree (%) (test set 5%)[c]	20.59	19.00	18.10	20.69	19.27	20.12
Protein atoms	2,250	2,250	2,250	2,250	2,250	2,250
Specific ligand atoms	0	30	56	42	45	45
Nonspecific ligand atoms	6	6	6	6	6	6
Water molecules	212	226	344	252	266	240
r.m.s.d. on bond lengths (Å)[d]	0.005	0.004	0.004	0.006	0.005	0.005
r.m.s.d. on bond angles (Å)[d]	0.639	0.581	0.620	0.598	0.725	0.518
Planar groups (Å)[d]	0.017	0.022	0.023	0.024	0.023	0.014
Chiral volume dev. ($Å^3$)[d]	0.045	0.044	0.037	0.040	0.047	0.024
Average B factor ($Å^2$)	28.53	17.71	18.90	18.52	23.04	24.85
Protein atoms	27.88	16.16	16.37	16.39	21.20	23.67
Specific ligand atoms	—	11.46	15.98	22.99	30.24	33.68
Nonspecific ligand atoms	46.34	40.25	38.58	44.04	56.73	47.14
Solvent atoms	34.90	33.33	35.54	36.14	36.63	33.77
ESRF beamline	ID29	ID29	ID23	ID23	ID23	ID23
Date of data collection	14 Mar 2015	14 Mar 2015	27 Nov 2015	27 Nov 2015	27 Nov 2015	27 Nov 2015
PDB code	6STM	6STN	6STO	2STP	6STQ	6STR

The highest-resolution shells used in the refinements are the following: crystal form1, 1.74–1.70 Å; 2, 1.80–1.75 Å; 3, 1.54–1.50 Å; 4, 1.64–1.60 Å; 5, 1.54–1.50 Å and 6, 1.85–1.80 Å. The nonspecific ligand is glycerol. The values in parentheses refer to the highest resolution shells. For the data collection of the different crystal forms (identified with a number in the second line) they are the following: 1, 1.74–1.70 Å; 2, 1.84–1.75 Å; 3, 1.58–1.50 Å; 4, 1.69–1.60 Å; 5, 1.58–1.50 Å and 6, 1.90–1.80 Å. [a]Rmerge = $\Sigma h \Sigma i$ | Iih – $<Ih>/\Sigma h \Sigma i < Ih >$, where $<Ih>$ is the mean intensity of the i observations of reflection h. [b]Rcryst = Σ | |Fobs| - |Fcalc| |/Σ |Fobs|, where |Fobs| and |Fcalc| are the observed and calculated structure factor amplitudes, respectively. Summation includes all reflections used in the refinement. [c]Rfree = Σ | |Fobs| - |Fcalc| |/Σ |Fobs|, evaluated for a randomly chosen subset of 5% of the diffraction data not included in the refinement. [d]r.m.s.d. from ideal values.

represented in the figure are the following A-1 1–43 [blue], A-2 44–86 [green] A-3 87–129 [orange] and A-4 130–170 [red]. The models were superimposed using the program LSQKAB (Kabsch 1978), and the mean r.m.s.d. values for equivalent α-carbons of domain A-1 and the other three were 0.97, 0.96 and 0.56 Å.

Structure of the dimer

The first line of evidence that indicated that ADL is a dimer under normal conditions was the appearance of a single band corresponding to the molecular weight of the dimer in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS PAGE) when the sample did not contain β-mercaptoethanol and a single band with the molecular weight of the protomer if the sample contained a reducing agent for disulfide bonds (data not shown). This experiment indicated that the dimer was stable even in the presence of SDS and that exposure to a disulfide bridge-reducing agent was required to dissociate it into the monomeric units. The two disulfide bridges linking the protomers in the dimer were easily identified since the quality of the electron density maps in the two areas was very good. The two bridges are established between Cys 42 of one chain and Cys 128 of the other.

Dynamic light scattering and mass spectrometry further confirmed the presence of a dimer in solution. Figure 4A is a scattered intensity distribution plot of different concentrations of ADL showing that the hydrodynamic diameter of the major peak of 7.1 nm is consistent with the presence of a dimer and that the samples contain a relatively low percentage of higher molecular weight aggregates. In a similar plot of the percent volume or number of particles, the second peak becomes totally insignificant. A mass spectrum of the protein sample (Figure 4B) shows the two peaks corresponding to the masses

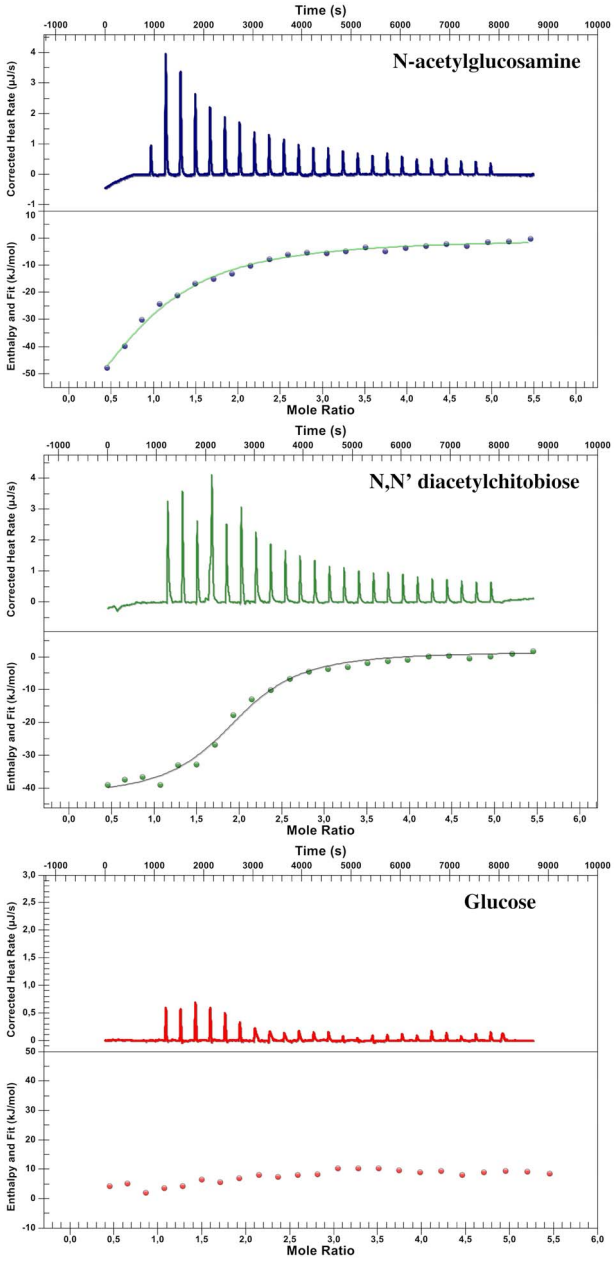

Fig. 2. Interaction of ADL with carbohydrates in solution. ITC titration of ADL with *N*-acetylglucosamine, *N,N′* diacetylchitobiose and glucose. The upper panels represent the raw data for the first injections (positive ordinate indicates exothermic), and the lower panels plot the corrected integrated heats as a function of the molar ratio. The best fit was obtained with an independent model for *N*-acetylglucosamine and *N,N′* diacetylchitobiose. The thermodynamic parameters are reported in the supplementary materials section (Supplementary Table SI). Note the absence of a significant heat of interaction of ADL with glucose.

of the protomer (16,869 m/z) and the dimer (33,562 m/z), which is clearly significantly higher.

Supplementary Table SII lists selected distances shorter than 3.5 Å which were measured between atoms in the interacting protomers. The total number of contacts below this cutoff is 112, but the table lists only one per interacting amino acid, the shortest. Table II is a reduced version of the table in the supplementary materials in which the cutoff is 2.8 Å.

Not surprisingly, the shortest contacts between the two protomers in the dimer are the two disulfide bridges established between Cys 42 of one chain and Cys 128 of the other in which the distance between the two SG atoms in the bridge is about 2 Å. A symmetric contact indicated in green in the table is also present between the NH2 of Arg 85 of one chain and the OD1 of Asp 87 of the other.

A representation of the dimeric molecule with the 2-fold axis in the plane of the picture and vertical (left) in two orientations differing from one another by a rotation of about 90° is shown in Figure 4C. The two protomers are represented in yellow and blue.

We have also calculated the solvent accessible area of protomers and dimer in crystal form 3, which diffracts to the highest resolution, 1.50 Å. The A monomer of ADL has a solvent accessible area of 7,225.9 Å^2 and the B monomer as a solvent accessible area of 7,317.1 Å^2. The contact areas are 2,118.4 and 2,142.8 Å^2, i.e. about 29% of the total surface of the monomers and a value that justifies the observed stability of this molecule when compared with those reported for other physiologically relevant dimers (Collaborative Computational Project Number 4 1994; Jones and Thornton 1995; Ponstingl et al. 2000).

The close contacts between the two protomers in the dimer play an important role in the function of the lectin in which binding of the carbohydrates to the protein involves residues from the two chains present in the dimer.

Interaction of the lectin with selected carbohydrates

Using the X-ray diffraction of single crystals, we have also studied the interaction of the lectin with *N*-acetylglucosamine, *N*-acetyllactosamine, *N*-acetylneuraminic acid and *N, N′* diacetylchitobiose.

The ADL dimer has two primary sugar-binding sites located at the interface of the two protomers with a participation in the contacts of amino acids from the two chains. We have labeled the primary site closest to the N-terminus of chain A as α and that closest to the N-terminus of chain B as β (Figure 5A). A secondary site is also present in the co-crystals with *N*-acetyllactosamine, *N*-acetylneuraminic acid and *N,N′* diacetylchitobiose (Figure 5C). Only the secondary site closest to the N-terminus of protomer B shows clear electron density for the carbohydrates since the equivalent site in protomer A is blocked by lattice contacts in the crystals, and our co-crystals were prepared by soaking of the apo-protein in solutions containing high concentrations of the sugars. We have labeled that third binding site as γ.

Table I summarizes the data collection and structure refinement statistics of the five different co-crystal forms. Crystal forms 5 and 6 contain both an excess of *N, N′* diacetylchitobiose, but the soaking times before exposure were different: 30 s in the first case and 1 min in the second.

The lectin binds *N*-acetylglucosamine in a very similar fashion in the two primary binding sites. Table III lists the shortest distances (cutoff = 3.35 Å) between ADL and *N*-acetylglucosamine. The residues that play equal roles in the two binding sites are on backgrounds and are colored in the same way. From the table, it is evident that the two polypeptide chains are equally important in providing the residues that participate in the contacts with the carbohydrate. The interactions of Glu 73 with the N2 of the monosaccharide explains why N-acetylation is essential for ligand-binding while binding to Asp 87 of the other chain, and a second

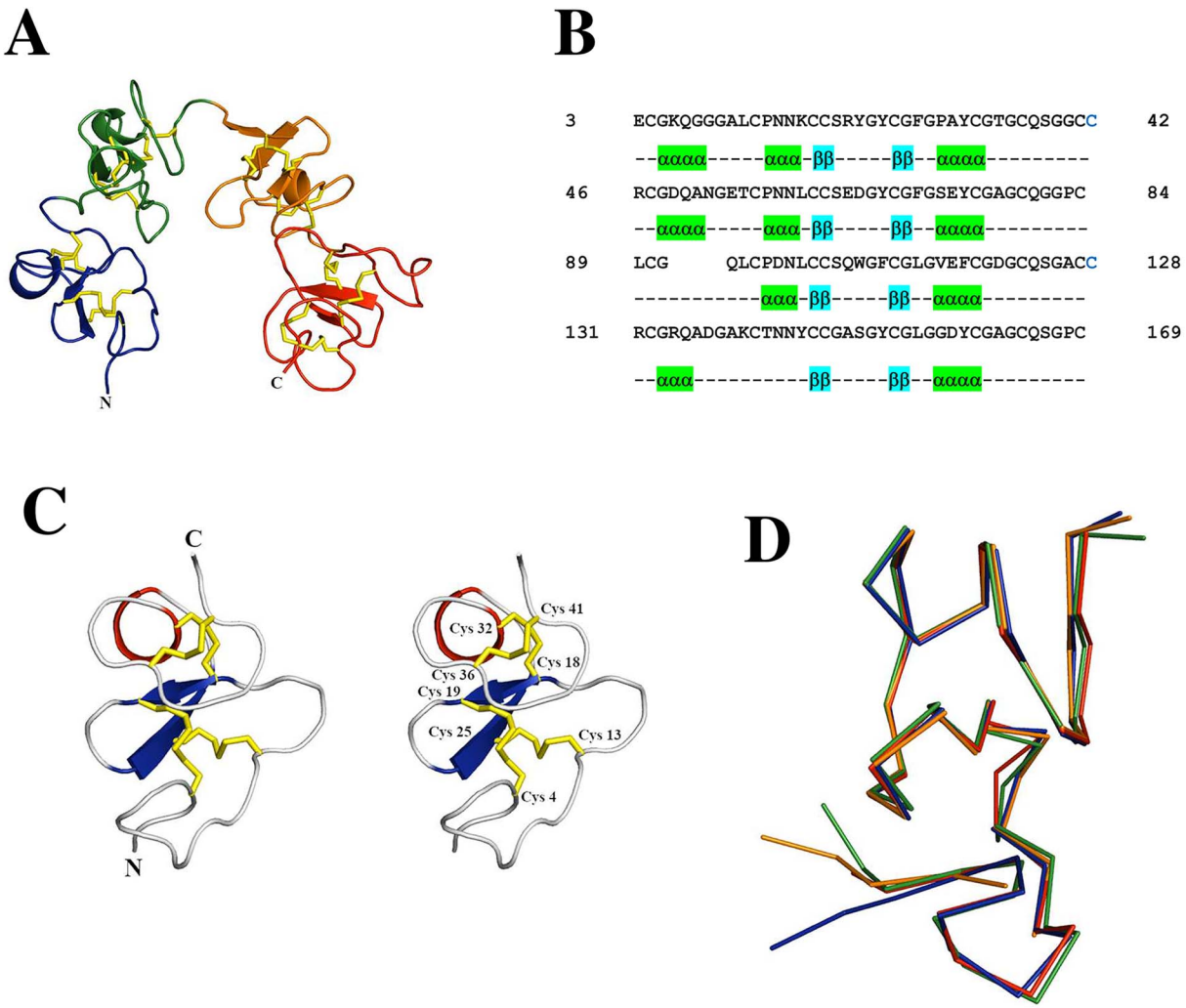

Fig. 3. X-ray structure of ADL. (**A**) Ribbon representation of the protomer in which each of the four domains has been colored differently. The coordinates used are those of monomer A of crystal form 1, and the four domains are represented in the following colors, starting from the N-terminus: A-1, blue; A-2, green; A-3, orange and A-4, red. The figure was prepared using the program PyMOL (http://www.pymol.org) that selected automatically the secondary structure. (**B**) Secondary structure assignments in the four polypeptide domains in a protomer. The four partial sequences were aligned on the basis of the positions of the disulfide bridges (not indicated in this figure), and the peptides superimposed span the chains 3–42, 46–84, 89–128 and 131–169. The amino acids in alpha helical conformation are green and those participating in the antiparallel beta sheets blue. The gap after amino acid 91 and before amino acid 97 corresponds to the disordered area in the electron density maps. (**C**) Stereo representation of the N-terminal domain A-1, spanning amino acids 1–43. The program PyMOL defined the secondary structure. (**D**) Superposition of the models of the four domains of protomer A. The amino acids represented in the figure are the following: A-1 1–43 [blue], A-2 44–86 [green] A-3 87–129 [orange] and A-4 130–169 [red]. The N-termini are in the figure at the bottom left. The models were superimposed using the program LSQKAB (Kabsch 1978).

contact with Ser 72 through the O4 of the sugar explains why *N*-acetylgalactosamine does not bind to the lectin. These two contacts with O4 confer a high selectivity to the lectin that does not bind *N*-acetylgalactosamine, which differs from *N*-acetylglucosamine only in the conformation of this epimeric oxygen.

The same distance cutoff was selected in Table IV that lists the shortest distances for *N*-acetylglucosamine, *N*-acetyllactosamine, *N*-acetylneuraminic acid and *N*,*N*′ diacetylchitobiose in the primary binding site, β. The interactions and distances are very similar for the first two and the last saccharides and are slightly different for *N*-acetylneuraminic acid in which, however, the main interactions of the acetylated N5 of *N*-acetylneuraminic acid are also established with Glu 73 of chain B. In this case, there is also an important contact between the OG of Ser 106 and O10 of *N*-acetylneuraminic acid and the O7 of the ligand with Trp 108 and Phe 110 of the protein.

A cutoff of 3.55 Å was also chosen for Table V to limit the number of distances in the table that are present in the secondary binding site, γ. In this case, most of the protein amino acids involved belong to chain B in the area close to the N-terminal portion of the chain, but Asp 158 and glycines 156 and 157 from chain A also play an important role in the interactions with *N*-acetyllactosamine and *N*-acetylneuraminic acid. The other residues from chain B that participate in these contacts are Ser 20, Tyr 24 and Tyr 31.

A stereo-diagram showing the electron density of *N*-acetylglucosamine in the primary binding sites of ADL with the two Fobs–Fcalc map contoured at a 1.5 σ level is shown in Figure 5A, and in Figure 5B, the main interactions observed between

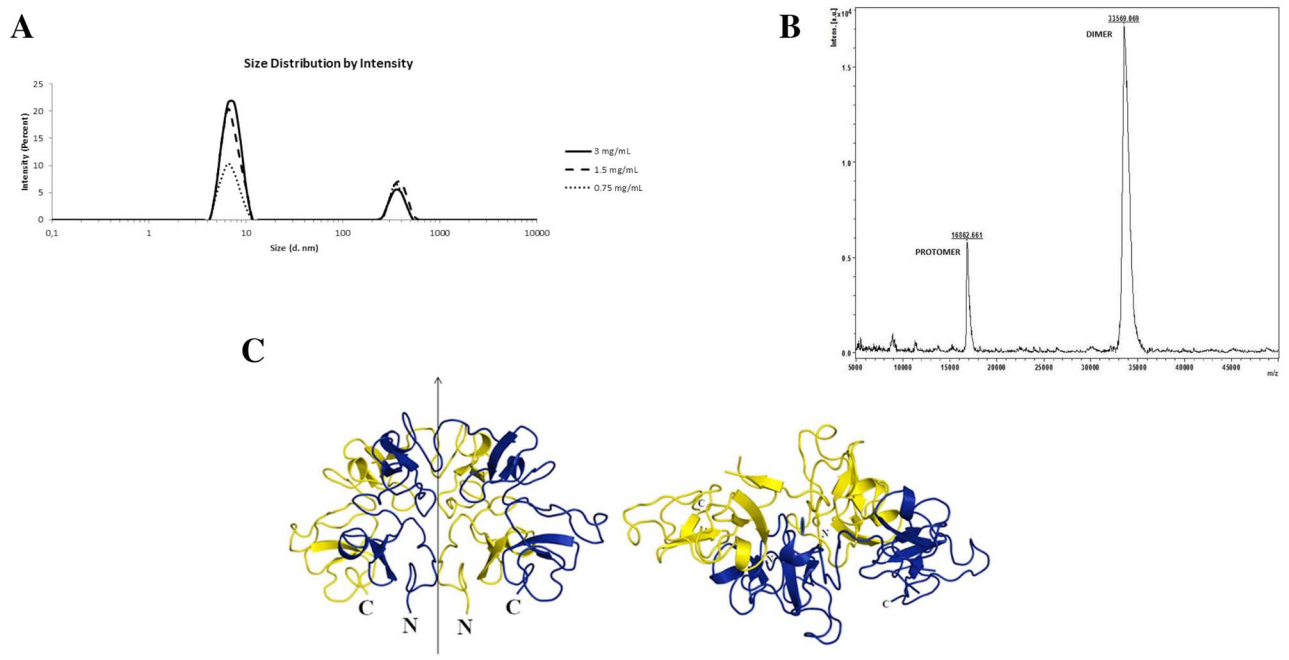

Fig. 4. The ADL dimer. (**A**) Dynamic light scattering experiment of ADL. The size distribution is by intensity, and the three curves corresponding to three different concentrations of the protein were recorded. The small high molecular weight peak on the right of the figure is due to the presence of a low percentage of aggregates. The hydrodynamic diameter of the highest peak corresponds to a radius of gyration of 7.1 nm, which is consistent with the presence of a dimer in solution and is in very good agreement with the value measured in the X-ray model. (**B**) MALDI spectrum of ADL showing the peaks corresponding to the molecular mass of the protomer and the dimer. (**C**) Ribbon representation of the ADL dimer. The two protomers are colored yellow and blue. The 2-fold axis present in the dimer is, on the left figure, in the plane and approximately vertical. The figure on the right has the 2-fold axis rotated approximately 90° and perpendicular to the plane.

Table II. Most significant contacts between the two protomers of ADL in the dimer in crystal form 1

Amino acid residue Protomer A	Atom	Distance (Å)	Amino acid residue Protomer B	Atom
Asn 16	OD1	2.66	Asn 59	ND2
Cys 42	SG	2.03	Cys 128	SG
Arg 85	NH2	2.76	Asp 87	OD1
Asp 87	OD1	2.71	Arg 85	NH2
Glu 116	OE1	2.35	Tyr 74	OH
Cys 128	SG	2.02	Cys 42	SG

The table lists only the shortest contacts (below 2.80 Å) between the amino acids in the two protomers. In yellow, the interchain disulfide bridges and, in green, a symmetric salt bridge.

N-acetylglucosamine and the primary binding site α of ADL are represented.

In Figure 5C, the electron density of *N*-acetyllactosamine which, as shown in the picture, is only clear for the *N*-acetylglucosamine moiety of the disaccharide, is represented in the two primary and in the secondary binding sites in different colors, and Figure 5D shows the interactions of the carbohydrate in the γ binding site of ADL.

Figure 5E is a ribbon representation of an ADL dimer with the electron density of the four saccharides, which are included in Table I. The ligand-binding site shown is β, and the ball and stick models of the carbohydrates are oriented as when bound to the lectin model represented in the figure. No important conformational changes are observed in the lectin when comparing its co-crystals with carbohydrates with those of the ligand-free protein.

Discussion

Exploiting its affinity for chitin, we have isolated a very abundant new lectin from the rhizomes of the very extensively diffused tall perennial grass, *A. donax*. On the average, we have purified approximately between 100 and 200 mg of protein, starting with 150 g of rhizomes. The protein is cataloged in the UniProt server under the name "uncharacterized protein" and is labeled as A0A0A9P802. The lectin was structurally characterized, i.e. its amino acid sequence and 3D structure were determined. Six different isoforms of the lectin could be separated by preparative isoelectric focusing, indicating a certain variability in the protein sequence. Only two of them were found to crystallize, and in the end, all the data for the final refinement of the six crystal forms, listed in Table I, were collected

Table III. Selected distances between the closest ADL residues and the *N*-acetylglucosamine molecule in the two primary binding sites; the cutoff is 3.35 Å

ADL residue	Atom	N-acetylglucosamine Atom	Distance (Å)
Ser 72 A	OG	C3	3.35 [α]
Ser 72 A	OG	O4	2.72 [α]
Ser 72 A	OG	O3	3.02 [α]
Glu 73 A	OE2	N2	2.90 [α]
Glu 73 A	OE1	O3	2.71 [α]
Asp 87 B	OD2	O4	2.85 [α]
Asp 87 B	OD1	O4	3.03 [α]
Asp 87 B	CG	O4	3.11 [α]
Asp 87 B	OD2	C6	3.24 [α]
Ser 106 B	OG	O7	2.52 [α]
Phe 110 B	CE2	O5	3.19 [α]
Phe 110 B	CD2	O5	3.34 [α]
Phe 110 A	CE2	O5	3.35 [β]
Ser 106 A	CB	O7	3.30 [β]
Ser 106 A	OG	O7	2.52 [β]
Asp 87 A	OD2	C4	3.29 [β]
Asp 87 A	CG	O4	3.24 [β]
Asp 87 A	OD1	O4	3.27 [β]
Asp 87 A	OD2	O4	2.60 [β]
Glu 73 B	OE2	N2	2.87 [β]
Glu 73 B	OE1	O3	2.87 [β]
Ser 72 B	OG	O3	3.13 [β]
Ser 72 B	OG	O4	2.68 [β]

The equivalent residues of the two chains are highlighted with the same color.

from only one isoform. The sequence of the refined models is 85.1% identical and 92.5% similar to that of protein A0A0A9P802 and, interestingly, has all the cysteines conserved with the exception of Cys 128 that forms the inter-protomer disulfide bridges with Cys 42 in the dimer (see Figure 1). We did not find any evidence of glycosylation of the side chains of the protein.

The molecular mass and amino acid sequence were also checked by mass spectrometric analyses using standard methods. Both in-source decay (ISD) analyses and peptide mass fingerprinting (PMF) were carried out. ISD coupled with matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) is used for the fast and reliable assignment of the protein chain termini. ISD is a fragmentation occurring in the MALDI source, after the laser shot and before the ion extraction. MALDI-ISD has been shown to efficiently produce *c*- and $z + 2$-type ions for the rapid analysis of both protein termini. When applied to intact ADL, the first 20 amino acids of the polypeptide chain were clearly identified and were found to be in agreement with the crystallographic result. PMF required the digestion of the protein with a specific endoprotease (trypsin) and the separation of the obtained fragments by gel electrophoresis, followed by the determination of their molecular masses. The list of the peptides analyzed and more details of the mass spectrometry experiments are given in the supplementary materials section (Supplementary Table SII).

The result of the X-ray analysis of single crystals of ADL revealed that it belongs to the well-known structural protein family of the hevein group of chitin-binding lectins (Itakura et al. 2017), which includes the extensively studied WGAs (Wright 1989). The carbohydrate-binding module, present in four copies in each polypeptide chain, has been named the hevein fold (Martínez-Caballero et al. 2014) and is also present in several plant chitinases (Kezuka et al. 2010).

ADL is a very steadfast dimer stabilized by two disulfide bridges formed between Cys 42 of one chain and Cys 128 of the other, a series of hydrophobic contacts and two symmetric hydrogen bonds established by the NH2 group of Arg 85 of one chain and the carboxylate of Asp 87 of the other. Dimerization of the molecule is undoubtedly relevant for its function since the four ligand-binding sites identified in the lectin require the participation of residues from both polypeptide chains in ligand-binding.

Binding of *N*-acetylglucosamine in the primary binding sites is mediated in ADL by several hydrophobic contacts mainly with Phe 110 and Phe 117 of chain B in site α and chain A in site β (Table SIV). Ligand specificity is due to the hydrogen bonds established by its O4 and the OG of Ser 72 and the ODs of Asp 87. In the *N*-acetyl moiety of the sugar molecule, there is a significant interaction between O7 of the carbohydrate and the OG of Ser 106. In both cases, the participation of residues from both protomers is equally important

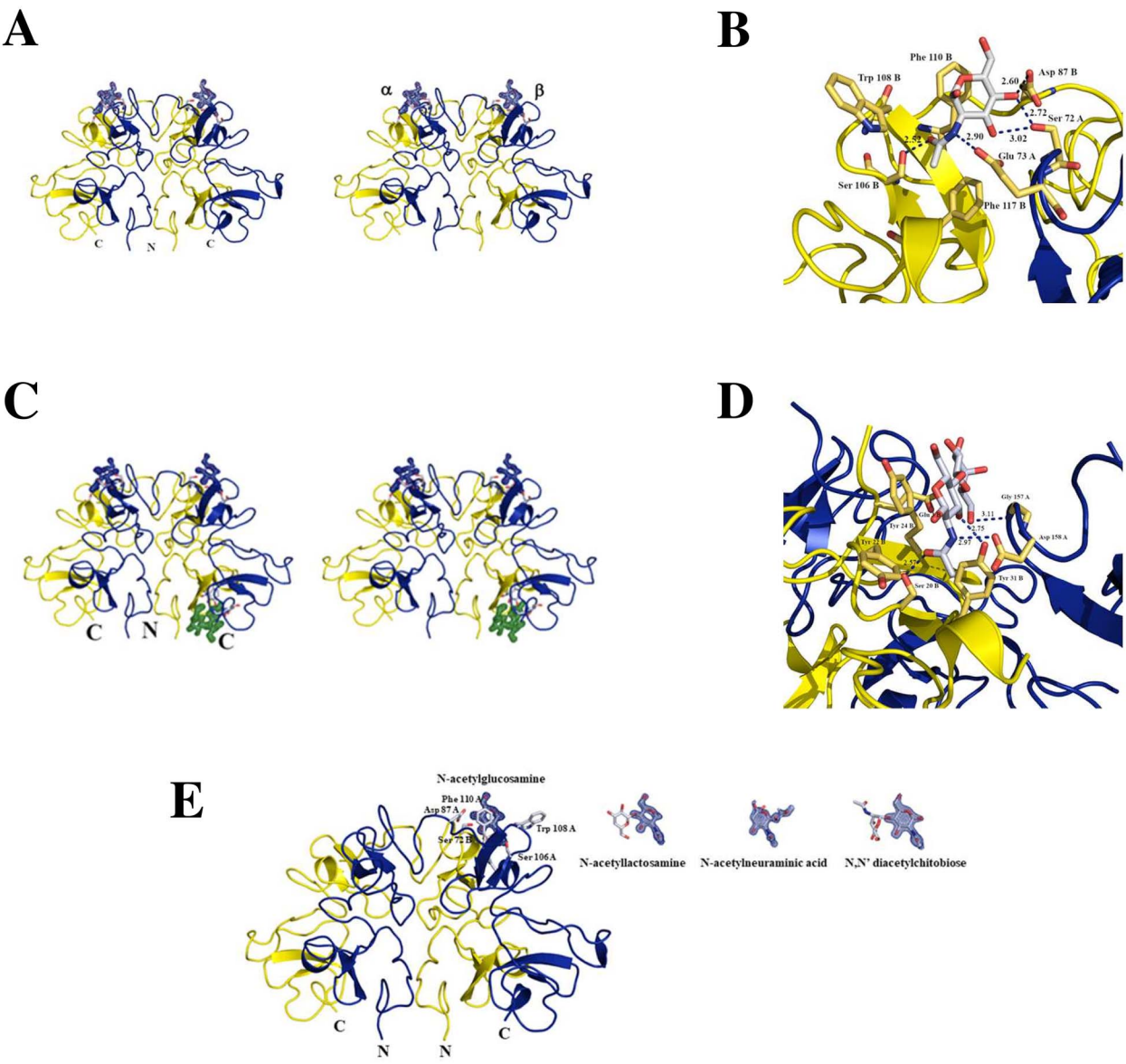

Fig. 5. Binding of carbohydrates to ADL. (**A**) Stereo diagram of an ADL dimer with two molecules of *N*-acetylglucosamine bound at the α and β primary sites. The electron density of the two Fobs-Fc maps was contoured at the 1.5 σ level. The side chains of the main amino acids involved in the interactions are represented in the figure. (**B**) Residues involved in the interaction of ADL with *N*-acetylglucosamine in binding site α. The most relevant distances are indicated in the figure. (**C**) Stereo diagram of an ADL dimer with the electron density of two molecules of *N*-acetyllactosamine bound at the α and β primary sites (blue) and a third molecule bound at the γ site (green). Note that in the α and β sites, the electron density observed corresponds to only the *N*-acetylglucosamine moiety of the disaccharide. The electron density of the two Fobs-Fc maps was contoured at the 1.5 σ level. (**D**) Main residues involved in the interaction of ADL with *N*,*N*′ diacetyllactosamine at the γ binding site (crystal form 3). (**E**) Ribbon representation of the ADL dimer with the Fobs-Fc electron density of the different saccharides that can occupy binding site β. The four carbohydrates represented as ball and stick models are *N*-acetylglucosamine, *N*-acetyllactosamine, *N*-acetylneuraminic acid and the *N*,*N*′ diacetylchitobiose. The electron density of the Fobs-Fc maps was contoured at a 1.5 σ level. The side chains of the main amino acids involved in the interactions are represented in the figure. The figure was prepared using the program PyMOL (http://www.pymol.org).

in ligand-binding (see Table III). The contacts with the epimeric O4 explain why *N*-acetylgalactosamine does not bind to this lectin.

The more significant contacts in this site are established with the *N*-acetylglucosamine moiety of the carbohydrates and in the two disaccharides studied, *N*-acetyllactosamine and *N*,*N*′ diacetylchitobiose, the second half of the carbohydrate molecule appears to be very flexible and there is no visible electron density in the maps (see Figure 5E).

Only one of the two available secondary sites was found occupied in our crystals due to the fact that in this particular crystal form the second site is blocked by intermolecular lattice interactions. The occupied site was labeled as γ, and the residues that interact with the ligands belong mainly to chain B but there is also a residue from chain A, Asp 158, that plays an essential role in conferring specificity to ligand-binding because it interacts with the N2 of the *N*-acetyl groups of the disaccharides and *N*-acetylneuraminic acid.

Table IV. Selected distances between the closest ADL residues of binding site β and the different carbohydrates

ADL residue	Atom	Distance (Å) [N-acetylglucosamine Atom]	Distance (Å) [N-acetyllactosamine Atom]	Distance (Å) [N-acetylneuraminic acid Atom]	Distance (Å) [N,N' diacetylchitobiose Atom]
Asp 87 A	OD2	3.29 [C4]	3.44 [C4]	-	3.39 [C4]
Asp 87 A	OD1	3.27 [O4]	3.34 [O4]	-	3.19 [O4]
Asp 87 A	OD2	2.60 [O4]	2.75 [O4]	-	2.67 [O4]
Asp 87 A	CG	3.24 [O4]	3.36 [O4]	-	3.25 [O4]
Ser 106 A	CB	3.30 [O7]	3.32 [O7]	-	3.35 [O7]
Ser 106 A	OG	2.52 [O7]	2.51 [O7]	-	2.59 [O7]
Ser 106 A	OG	-	-	2.69 [O10]	-
Phe 110 A	CE2	3.35 [O5]	3.31 [O5]	-	3.51 [O5]
Glu 116 A	O	3.42 [C8]	3.36 [C8]	-	3.33 [C8]
Glu 116 A	O	-	-	3.21 [C11]	-
Glu 73 B	OE2	-	-	2.71 [N5]	-
Glu 73 B	OE1	-	-	3.31 [C4]	-
Glu 73 B	OE1	-	-	2.86 [O4]	-
Glu 73 B	OE2	2.87 [N2]	2.87 [N2]	-	2.79 [N2]
Glu 73 B	OE1	2.87 [O3]	2.83 [O3]	-	2.91 [O3]
Ser 72 B	OG	2.68 [O4]	2.63 [O4]	-	2.58 [O4]
Ser 72 B	OG	3.40 [C3]	3.39 [C3]	-	3.25 [C3]
Ser 72 B	OG	3.13 [O3]	3.13 [O3]	-	3.11 [O3]
Ser 72 B	CB	3.41 [O4]	3.37 [O4]	3.77 [O4]	3.27 [O4]

The distances refer to site β of the asymmetric unit of crystal forms' numbers 2, 3, 4 and 6. The maximum distance cutoff is 3.35 Å for at least one of the carbohydrates. In colored background, the shortest distance of the residue in contact is given.

Chitin, the long chain polymer recognized by ADL is the second most abundant carbohydrate found in nature and a major component of the cell wall of fungi, the cell membrane of yeast and bacteria and of the exoskeleton of crustaceans and insects. Since plants do not contain chitin but most plant pathogens do, this polysaccharide plays an important role in triggering the plant innate immune system against insects, yeast, bacteria and fungi (De Hoff et al. 2009; Kombrink et al. 2011). The plant response to aggression by chitin-containing pathogens mobilizes a series of molecules of the defense signaling pathways in which the hevein-like proteins are believed to play a role. Therefore, it is likely that ADL may be a member of the defense mechanism of the plant against insects, yeast, fungi and bacteria (Berthelot et al. 2016).

When the coordinates of ADL are analyzed by the Dali server (Holm and Sander 1999), WGA, isoform 3, is chosen by the program as the protein whose structure is most similar to the queried molecule. The r.m.s.d. for the alpha carbons of ADL and WGA-3 is 3.692 Å for monomer A and 3.589 Å for monomer B of ADL and the equivalent chains of WGA-3.

Figure 6A compares the amino acid sequences of ADL, WGA-3 and WGA-1. The latter was included in the figure to give an idea of the sequence variability between the different isoforms of WGA and to allow a comparison of the residues that bind *N*-acetylglucosamine in site α of ADL and WGA. The experimental X-ray data on the binding of the carbohydrate to this site are available only for isoform 1 of WGA. As the figure shows, the three proteins show a high degree of sequence similarity and have all of their cysteine residues conserved with two highly significant exceptions, Cys 42 and 128, the two cysteines that form the disulfide bridges that link the two protomers in the ADL dimer and absent in the two isoforms of WGA.

In the figure, the conserved cysteines are represented on a red background, while the two that form the disulfide bridges of ADL are represented on a green background. The two models are drawn superimposed in Figure 6B, a stereo representation, in which the ADL protomers are yellow and blue and the WGA-3 chains (PDB code 1K7T; Harata et al. 1995) are red and green. Note that the two models superimpose quite well and that the relationship between the two protomers in the dimer is very similar. The molecular models of *N*-acetylglucosamine bound to the α primary site of ADL (left) and WGA-1 (right) (Schwefel et al. 2010; PDB accession code 2UVO) are represented in Figure 6C. A direct comparison with WGA-3 is not possible because its crystal structure with *N*-acetylglucosamine is not available, but the residues involved in the contacts are conserved in the two isoforms (see Figure 6A). Sugars and relevant residues are shown in sticks models and coordinate bonds as yellow dashed lines. Our ITC experiments yield thermodynamic values for the ΔH of interaction with both *N*-acetylglucosamine and *N*,*N*′ diacetylchitobiose, which are of the same order of magnitude as those observed for WGA (Bains et al. 1992). In every case, the values of $T\Delta S$ are negative indicating that the binding process is enthalpically driven.

Many important practical biotechnological applications have been proposed and implemented for the lectins of the WGA family. Given the similarities of ADL with this protein group and the ease of preparation, yield, cost and ready availability of the protein source, it can be anticipated that most of those uses may be carried out equally well by the protein we describe in this paper.

Materials and methods

Protein purification

ADL was purified from the rhizomes of the common giant cane by affinity chromatography in a column of chitin (New England Biolabs).

In each preparation, 200 g of rhizomes were homogenized in a blender using a volume of approximately 300 mL of Tris–HCl: 50 mM, pH: 7.5, NaCl: 0.5 M, NaN_3: 0.02% and EDTA: 1 mM. After centrifugation at 7,000 g for 30 min, the supernatant was

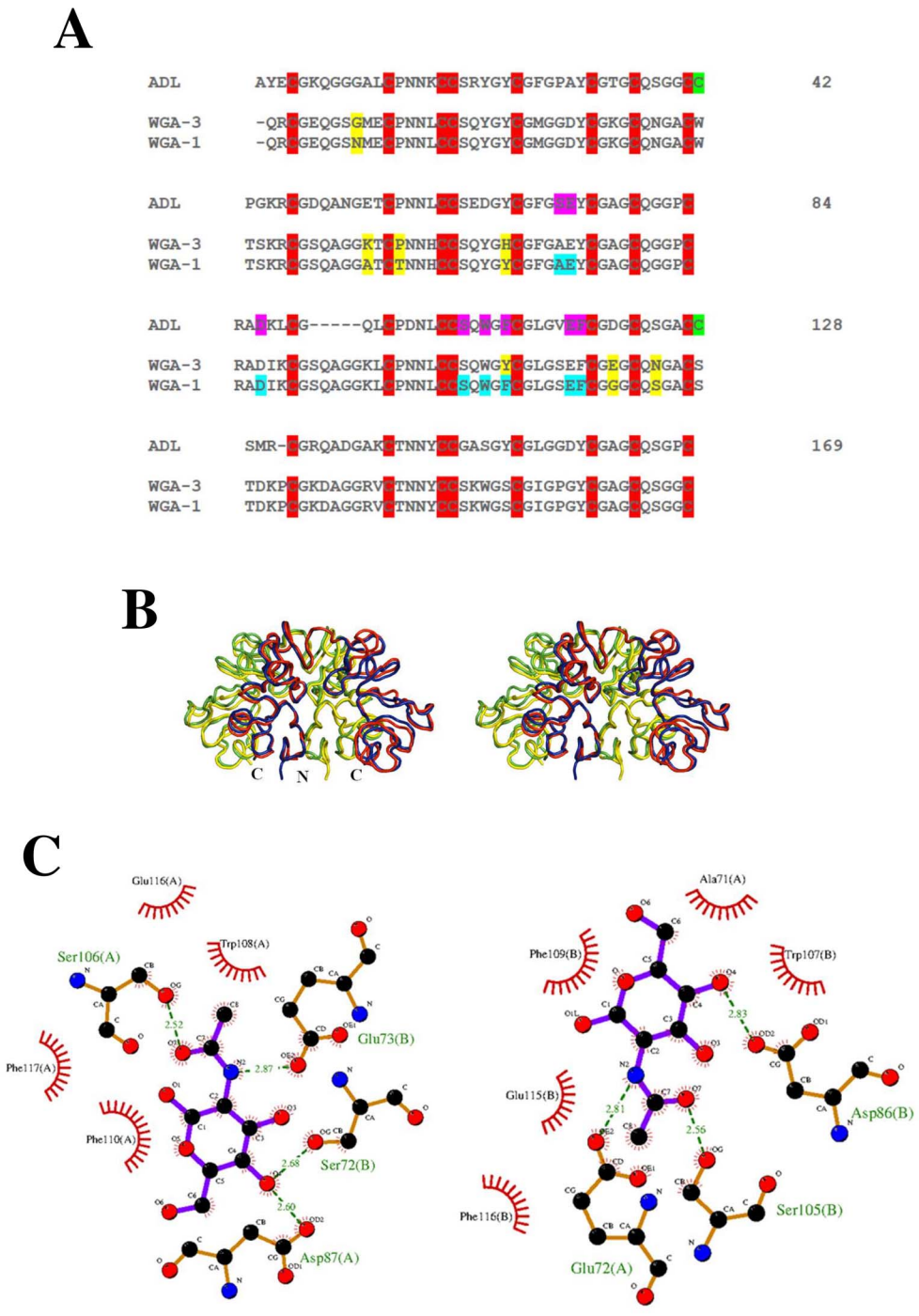

Fig. 6. Comparison of ADL and WGA. (**A**) Sequence comparison of ADL and WGA, isoform 3 (UniProtKB - P10969) and isoform 1 (UniProtKB - P10968). The conserved cysteines are indicated on a red background, while the two nonconserved (numbers 42 and 128) are highlighted on a green background. The residues that are different in the two WGA isoforms are highlighted on a yellow background. The magenta residues in ADL are those involved in ligand-binding site α and those in blue are those involved in the equivalent site of WGA-1 (PDB code 2UVO). (**B**) Stereo representation of the ADL and WGA, isoform 3 (WGA-3, Harata et al. 1995, PDB accession code 1WGT), dimers superimposed. The two protomers of ADL are yellow and blue whereas those of WGA-3 are red and green. The superposition of the two dimers was carried out using the program LSQKAB (Kabsch 1978) and shows clearly that the contacts between the protomers in the two dimers are very similar. (**C**) Molecular models of *N*-acetylglucosamine bound to the α primary site of ADL (left) and WGA-1 (right) (Schwefel et al. 2010, PDB accession code 2UVO). A direct comparison with WGA-3 is not possible because its crystal structure with *N*-acetylglucosamine is not available, but the residues involved in the contacts are conserved in the two isoforms (**A**). Sugars and relevant residues are shown in sticks models and coordinate bonds as green dashed lines. The figure was prepared using the program LigPLOT^{+} (Laskowski and Swindells 2011).

Table V. Selected distances between the closest ADL residues of binding site γ and the different carbohydrates

ADL residue	Atom	Distance (Å) [N-acetyllactosamine Atom]	Distance (Å) [N-acetylneuraminic acid Atom]	Distance (Å) [N,N' diacetylchitobiose Atom]
Glu 3 B	OE1	3.33 [C5]	-	-
Glu 3 B	OE1	3.30 [C6]	-	-
Ser 20 B	CB	3.32 [OAZ]	-	-
Ser 20 B	OG	2.57 [OAZ]	-	-
Ser 20 B	CB	-	-	3.20 [O7]
Ser 20 B	OG	-	-	2.51 [O7]
Ser 20 B	OG	-	2.66 [O10]	-
Tyr 24 B	CB	-	-	3.16 [O3]
Cys 25 B	O	2.78 [O6]	-	-
Tyr 31 B	OH	3.31 [NAT]	-	-
Tyr 31 B	OH	2.75 [OAV]	-	-
Tyr 31 B	OH	-	-	2.66 [C3]
Tyr 31 B	CE2	-	-	2.35 [O3]
Tyr 31 B	CZ	-	-	2.58 [O3]
Tyr 31 B	OH	-	-	2.16 [O3]
Tyr 31 B	OH	-	2.60 [O4]	
Tyr 31 B	CZ	-	3.29 [C11]	-
Tyr 31 B	OH		3.27 [N5]	-
Asp 158 B	OD2	-	-	3.04 [C8]
Asp 158 B	OD1	-	-	3.08 [N2]
Gly 156 A	CA	3.32 [O6]	-	-
Gly 157 A	N	3.11 [O6]	-	-
Asp 158 A	OD2	-	3.23 [C11]	-

The distances refer to site γ, closest to the N-terminus of protomer B of the asymmetric unit of crystal forms' numbers 3, 4 and 6. The equivalent site closest to the N-terminus of chain A is not occupied in the crystals and the maximum distance cutoff is 3.35 Å.

filtered through glass wool to eliminate the solid aggregates. The extraction was repeated twice and, after combining the three extracts, the total volume was concentrated to about 100 mL under nitrogen pressure in an Amicon (New England Biolabs, Ipswich, MA USA) cell with a polyethersulfone membrane with 10 kD cutoff. The concentrated extracts were loaded onto a column (2.5 × 15 cm) containing about 10 mL of the resin, and the column was washed with 50 mL of Tris–HCl: 50 mM, pH: 7.5, NaCl: 0.5 M, NaN_3: 0.02%, EDTA: 1 mM and 0.25% of Triton X-100. After extensive washing with Tris–HCl: 20 mM, pH: 7.5, NaCl: 0.15 M, NaN_3: 0.02%, EDTA: 1 mM, and when the O.D. of the solution was negligible, the protein was eluted with the same buffer containing 0.4 M of *N*-acetylglucosamine.

The protein solution was subsequently submitted to a hydrophobic interaction chromatography step using a Lipidex 1000 column (1.5 × 20 cm containing about 10 mL of the resin) at 37°C to remove potential hydrophobic ligands bound to the lectin. The column was washed with the Tris buffer that did not contain *N*-acetylglucosamine.

After concentration to an adequately reduced volume of the sample, the protein solution was loaded onto a Superdex G75 gel filtration column (1.5 × 60 cm). Elution was carried out with the Tris buffer that did not contain *N*-acetylglucosamine. A single band was present in SDS PAGE, but six different isoforms were revealed by isoelectric focusing. They were separated by preparative isoelectric focusing using the Ampholyne carriers in the pH range: 7.0–9.0. The bands were eluted from the Ultrodex gel matrix with Tris–HCl: 20 mM, pH: 7.5, NaCl: 0.15 M, NaN_3: 0.02% and were used for the crystallization experiments. Only two of the six bands crystallized and both under the same conditions, either separated or together. The final yield was about 100 mg of all the forms for 200 g of the starting material.

Cloning and cDNA sequence

Total RNA was isolated from fresh *A. donax* rhizomes using the TRIZOL (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany) protocol. 1 µg of RNA was used to synthetize the first cDNA strand by means of the Superscript First Strand Synthesis System (Invitrogen, Waltham, USA) using an $Oligo(dT)_{18}$ primer. The ADL sequence was amplified by PCR using a High Fidelity Taq DNA polymerase (Jena Biosiences, Jena, Germany) with two primers based on the sequence of *A. donax* (giant reed) hypothetical protein present on the GenBank server with code number JAF18563.1. The primers used were the following: forward, ATGGCAATGTTCCTTGCAGCG, reverse, CTATGCAACCAAAGCGATGGA. The amplicon, separated on a 1% agarose gel, was eluted and directly TA-cloned into the pGEM-T Vector System (Promega). The plasmids from four positive colonies were isolated and sequenced using the standard T7 forward and SP6 reverse primers.

Calorimetric techniques

Calorimetric analyses were performed with a TA Nano ITC system. The Nano ITC was loaded with 200 µL of PBS as reference and with

200 µL of purified protein solution with a concentration of 50 µM in PBS in the sample cell. The sugar concentration in the syringe was 1 mM (in phosphate buffer pH 6.5) and the titration consisted of 25 injections of 2 µL every 10 min with a stirring speed of 250 rpm. Experiments were performed at 20°C in triplicate.

The resulting thermograms were analyzed and compared using the Nano Analyse Software.

MALDI-TOF MS analyses

The highly purified preparations of ADL were precipitated with acetone, then resuspended in a TFA solution (50:50 acetonitrile: water, 0.1% TFA) and incubated for 30 min at room temperature under shaking. The resulting samples were diluted 1:100 with the same solution, then rediluted with sinapinic acid (10 mg/mL in ACN:H20 1:1 with 0.1% TFA), and 1 µL of the sample/matrix solution was spotted in triplicate onto a Ground steel MALDI target plate (Bruker Daltonics, Billerica, Massachusetts, USA) and was allowed to dry at room temperature. MALDI-TOF MS analyses of intact proteins were performed on a Bruker UltrafleXtreme MALDI-TOF/TOF instrument (Bruker Daltonics, Billerica, Massachusetts, USA). Mass spectra were collected from m/z 5,000 to 50,000 in positive linear mode and mass calibration was performed using a protein standard mixture composed of bovine serum albumin, protein A and trypsinogen.

Mass spectra obtained under the described conditions show a peak which confirms the MW of the protein.

Then, ISD spectra were obtained in order to have sequence information directly from the intact protein. The spectra obtained confirm the N-terminal sequence of the protein.

Analysis of the peptides were performed in a reflectron mode in order to increase the mass accuracy, and mass spectra were analyzed in the range m/z 550–4,000 by averaging the data from 500 laser shots. The peaks obtained were analyzed by the Biotools Software (version 3.2, Bruker Daltonics), the amino acid sequence was simulated by the Expasy software and the presence of the expected peptides was confirmed by MS/MS.

The peaks observed in the MALDI-TOF analyses of the two isoforms studied are included in the supplementary materials section (Supplementary Figure S1).

Crystallization and X-ray structure determination

Crystals were grown using the vapor diffusion hanging drop method by mixing equal volumes (1 µL) of the protein solution at a concentration of 30 mg/mL and with a solution of 7.5% in PEG 8000, 10% ethylene glycol and 0.1 M sodium cacodylate, pH 6.0 as the precipitant. The crystals belong to the hexagonal space group $P6_5$ with the following unit cell parameters for the apo-protein: a = 127.0 Å and c = 46.5 Å, and contain an ADL dimer in the asymmetric unit. A summary of the data of all the different crystal forms is presented in Table I.

The diffraction data were collected from crystals frozen at 100°K after a brief immersion in a mixture of 80% of the mother liquor and 20% glycerol.

The crystal structure of the apo-protein was solved initially by the MIR method using platinum and mercury derivatives and collecting the data at the ESRF in Grenoble. The heavy atom derivatives were prepared by the overnight soaking of a crystal in mother liquor saturated with the addition of either K_2PtCl_4 or EMP (ethyl mercury phosphate). The three major platinum sites were located in a difference Patterson map (Sheldrick 2008) and were refined using the program MLPHARE (Collaborative Computational Project Number 4; Winn et al. 2011) and were subsequently used as input for the program autoSHARP (Bricogne et al. 2003) that was used to locate another Pt and two Hg sites. With these six sites, the anomalous data of the Pt derivative and density modification the data were phased.

The electron density map thus produced was of excellent quality and could be readily interpreted. The initial model of the apo-protein was built in the high quality map at 2.5 Å resolution using the program Coot (Emsley et al. 2010). Refinement was carried out using the program REFMAC (Murshudov et al. 1997). During the process of refinement and model building, the quality of the model was controlled with the program PROCHECK (Laskowski et al. 1993). The refined coordinates were used to identify WGA isoform 3 as the structurally most similar protein present in the Protein Data Bank, and the highest resolution available coordinates (PDB code 2X52, Schwefel et al. 2010) were used to resolve the structure of the apo-protein and of the co-crystals by molecular replacement using the program MOLREP (Vagin and Teplyakov 2000).

The data used for the final refinement of all the forms, apo-protein and co-crystals, were collected at the ID23 and ID29 beamlines of the European Synchrotron Radiation Facility in Grenoble. The data were indexed, integrated and reduced using the programs MOSFLM (Leslie 1992) and Scala (Collaborative Computational Project Number 4). The final refinement of the six crystal forms of ADL was carried out initially using the program REFMAC (Murshudov et al. 1997) and, in a second stage, with the program Phenix.refine (Adams et al. 2010). The models were built using the program Coot (Emsley et al. 2010) and were finally subjected to a final round of TLS refinement. The saccharides in the co-crystals were modeled into difference Fourier maps phased by the refined, unliganded structure. The models of the complexes were refined with REFMAC using the same criteria followed in the refinement of the apo-protein. Solvent molecules were added to the models in the final stages of refinement according to hydrogen-bond criteria and only if their B factors refined to reasonable values and if they improved the Rfree.

The diffraction data and refinement statistics of all the models are summarized in Table I. The figures of the models were prepared using the program PyMOL (http://www.pymol.org).

Supplementary data

Supplementary data are available at *Glycobiology* online.

Acknowledgements

We are grateful to the Mass Spectrometry Facility (Centro Piattaforme Tecnologiche) of the University of Verona for the use of the facility and to the staff of the ESRF Synchrotron in Grenoble (Proposal MX 1750) for their assistance during data collection. We also thank the Department of Biotechnology of the University of Verona for the fellowship (Assegno di Ricerca) received by F.P. The coordinates of the models and the structure factors of the apo-protein and the complexes with carbohydrates have been deposited in the protein data bank. The PDB accession codes are 6STM, 6STN, 6STO, 6STP, 6STQ and 6STR.

Funding

Fondazione Cassa di Risparmio di Verona, Vicenza, Belluno e Ancona.

Conflict of interest statement

None declared.

Abbreviations

ADL, *Arundo donax* lectin; cDNA, complementary deoxyribonucleic acid; DNA, deoxyribonucleic acid; MALDI-TOF MS, matrix-assisted laser desorption ionization time of flight mass spectrometry; PEG, polyethylene glycol; RNA, ribonucleic acid; SDS PAGE, sodium dodecylsulfate-polyacrylamide gel electrophoresis.

References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 66:213–221.

Bains G, Reiko TL, Yuan CL, Freire E. 1992. Microcalorimetric study of wheat germ agglutinin binding to *N*-acetylglucosamine and its oligomers. *Biochemistry*. 127:258–270.

Berthelot K, Peruch F, Lecomte S. 2016. Highlights on *Hevea brasiliensis* (pro)hevein proteins. *Biochimie*. 127:258–270.

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. 2004. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem J*. 382:769–781.

Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W. 2003. Generation, representation and flow of phase information in structure determination: Recent developments in and around SHARP 2.0. *Acta Crystallogr Sect D Biol Crystallogr*. 59:2023–2030.

Chauhan JS, Rao A, Raghava GPS. 2013. In silico platform for prediction of N-, O- and C-glycosites in eukaryotic protein sequences. *PLoS One*. 8:e67008.

Corno L, Pilu R, Adani F. 2014. *Arundo donax L.*: A non-food crop for bioenergy and bio-compound production. *Biotechnol Adv*. 32:1535–1549.

De Hoff PL, Brill LM, Hirsch AM. 2009. Plant lectins: The ties that bind in root symbiosis and plant defense. *Mol Genet Genomics*. 282:1–15.

Devi RV, Basil-Rose MR. 2018. Lectins as ligands for directing nanostructured systems. *Curr Drug Deliv*. 15:448–452.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 66:486–501.

Gonzàlez De Mejia E, Prisecaru VI. 2005. Lectins as bioactive plant proteins: A potential in cancer treatment. *Crit Rev Food Sci Nutr*. 45:425–445.

Harata K, Nagahora H, Jigami Y. 1995. X-ray structure of wheat germ agglutinin isolectin 3. *Acta Crystallogr SectD*. 51:1013–1019.

Holm L, Sander C. 1999. Protein folds and families: sequence and structure alignments. *Nucleic Acids Res*. 27:244–247.

Itakura Y, Nakamura-Tsuruta S, Kominami J, Tateno H, Hirabayashi J. 2017. Sugar-binding profiles of chitin-binding lectins from the hevein family: A comprehensive study. *Int J Mol Sci*. 30:18. pii: E1160.

Jiménez-Barbero J, Javier Cañada F, Asensio JL, Aboitiz N, Vidal P, Canales A, Groves P, Gabius HJ, Siebert HC. 2006. Hevein domains: An attractive model to study carbohydrate protein interactions at atomic resolution. *Adv Carbohydr Chem Biochem*. 60:303–354.

Jones S, Thornton JM. 1995. Protein-protein interactions: A review of protein dimer structures. *Prog Biophys Mol Biol*. 63:31–65.

Kabsch W. 1978. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr*. A32:922–923.

Kaur A, Singh J, Kamboj SS, Sexana AK, Pandita RM, Shamnugavel M. 2005. Isolation of an N-acetyl-D-glucosamine specific lectin from the rhizomes of *Arundo donax* with antiproliferative activity. *Phytochemistry*. 66:1933–1940.

Kezuka Y, Kojima M, Mizuno R, Suzuki K, Watanabe T, Nonaka T. 2010. Structure of full-length class I chitinase from rice revealed by X-ray crystallography and small-angle X-ray scattering. *Proteins*. 78: 2295–2305.

Kombrink A, Sánchez-Vallet A, Thomma BP. 2011. The role of chitin detection in plant–pathogen interactions. *Microbes Infect*. 13: 1168–1176.

Lannoo N, Van Damme EJ. 2014. Lectin domains at the frontiers of plant defense. *Front Plant Sci*. 5(397):1–16. doi: 10.3389/fpls.2014.00397.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst*. 26:283–291.

Laskowski RA, Swindells MB. 2011. LigPlot^{+}: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model*. 51(10): 2778–2786.

Leslie AGW. 1992. Recent changes to the MOSFLM package for processing film and image plate data. *Jnt CCP4/ESF-EACMB Newslett Protein Crystallogr*. 26:27–33.

Martínez-Caballero S, Cano-Sánchez P, Mares-Mejía I, Díaz-Sánchez AG, Macías-Rubalcava ML, Hermoso JA, Rodríguez-Romero A. 2014. Comparative study of two GH19 chitinase-like proteins from *Hevea brasiliensis*, one exhibiting a novel carbohydrate-binding domain. *FEBS J*. 281:4535–4554.

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 53:240–255.

Ponstingl H, Henrick K, Thornton JM. 2000. Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins*. 41: 47–57.

Santín-Montanyá MI, Jiménez J, Ocaña L, Sánchez FJ. 2013. Effects of sprout cutting plus systemic herbicide application on the initial growth of giant reed. *J Environ Sci Health B*. 48:285–290.

Schwefel D, Maierhofer C, Beck JG, Seeberger S, Diederichs K, Moller HM, Welte W, Wittmann V. 2010. Structural basis of multivalent binding to wheat germ agglutinin. *J Am Chem Soc*. 132:8704–8719.

Sharon N. 2007. Lectins: Carbohydrate-specific reagents and biological recognition molecules. *J Biol Chem*. 282:2753–2764.

Sheldrick GM. 2008. A short history of SHELX. *Acta Cryst A*. 64: 112–122.

Souza MA, Carvalho FC, Ruas LP, Ricci-Azavedo R, Roque-Barraira MC. 2013. The immunomodulatory effect of plant lectins: A review with emphasis on ArtinM properties. *Glyconj J*. 30:641–657.

Taylor, M.E. Drickamer, K. 2003 *Introduction to Glycobiology*, Vol. XV. Oxford, New York: Oxford University Press. p. 207.

Vagin A, Teplyakov A. 2000. An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr*. 56: 1622–1624.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, et al. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*. 67:235–242.

Wright CS. 1989. Comparison of the refined crystal structures of two wheat germ isolectins. *J Mol Biol*. 209:475–487.