Running head:

Binding properties of PP2-A1 phloem lectin

Author for correspondence:

Dr. Sylvie DINANT
Institut Jean-Pierre Bourgin, UMR1318 INRA-AgroParisTech
INRA Centre de Versailles, Route de Saint-Cyr, 78026 Versailles Cedex France

Tel: 33 (0) 1 30 83 030 47
E-mail: dinant@versailles.inra.fr
Title:

Binding properties of the N-acetylglucosamine and high-mannose N-glycan PP2-A1 phloem lectin in Arabidopsis

Authors' full names:

BENETEAU Julie\textsuperscript{1,2}, RENARD Denis\textsuperscript{2}, MARCHÉ Laurent\textsuperscript{2}, DOUVILLE Elise\textsuperscript{2}, LAVENANT Laurence\textsuperscript{2}, RAHBÉ Yvan\textsuperscript{3}, DUPONT Didier\textsuperscript{4}, VILAINÉ Françoise\textsuperscript{1,5}, DINANT Sylvie\textsuperscript{1}

Authors' addresses:

\textsuperscript{1}Institut Jean Pierre Bourgin (IJPB), UMR1318 INRA-AgroParisTech, INRA Centre de Versailles-Grignon, F-78000 Versailles, France
\textsuperscript{2}UR1268 Unité de Recherches Biopolymères Interactions Assemblages (URBIA), INRA F-44300 Nantes, France
\textsuperscript{3}UMR203 Laboratoire Biologie Fonctionnelle, Insectes et Interactions (BF2I), INSA INRA F-69621 Villeurbanne, France
\textsuperscript{4}UMR1253 Sciences et Technologies du Lait et de l’Oeuf (STLO), INRA F-35042 Rennes, France
\textsuperscript{5}Laboratoire Commun de Biochimie (LCB), UMR1318 INRA-AgroParisTech, INRA Centre de Versailles-Grignon, F-78000 Versailles, France
Footnotes:

Address(es) of authors if different from heading:
BENETEAU Julie; Unité d'Ecologie et de Physiologie du Système Digestif (UEPSD), INRA
Domaine de Vilvert 78352 JOUY-EN-JOSAS

Corresponding author with e-mail address:
DINANT Sylvie; E-mail: Sylvie.Dinant@versailles.inra.fr
Abstract

Phloem protein 2 (PP2) is a component of the P-protein bodies found in sieve elements. We describe here the lectin properties of the Arabidopsis PP2-A1. Using a recombinant protein produced in *E. coli*, we demonstrated binding to N-acetyl-glucosamine oligomers. Glycan array screening showed that PP2-A1 also bound to high-mannose N-glycans and 9-acyl-N-acetyl-neuraminic sialic acid. Fluorescence spectroscopy-based titration experiments revealed that PP2-A1 had two classes of binding site for N,N’,N”triacylchitotriose: a low-affinity and a high-affinity site, promoting the formation of protein dimers. A search for structural similarities revealed that PP2-A1 aligned with the Cbm4 and Cbm22-2 carbohydrate-binding modules, leading to the prediction of a β-strand structure for its conserved domain. We investigated whether PP2-A1 interacted with phloem sap glycoproteins, by first characterizing abundant Arabidopsis phloem sap proteins by LC MS/MS. Then we demonstrated that PP2-A1 bound to several phloem sap proteins, and that this binding was not completely abolished by glycosidase treatment. As many plant lectins have insecticidal activity, we also assessed the effect of PP2-A1 on weight gain and survival in aphids. Unlike other mannose-binding lectins, when added to an artificial diet, recombinant PP2-A1 had no insecticidal properties against *Acyrthosiphon pisum* and *Myzus persicae*. However, at mid-range concentrations, the protein affected weight gain in insect nymphs. These results indicate the presence in PP2-A1 of several carbohydrate-binding sites, with potentially different functions in the trafficking of endogenous proteins or in interactions with phloem-feeding insects.

Key words: Phloem Protein 2, *Myzus persicae, Acyrthosiphon pisum*, proteome, phloem sap, GlcNAc, carbohydrate binding protein (CBP)
Introduction

The vascular tissues of land plants are responsible for the allocation of nutrients from their sites of uptake and synthesis. They also act as major routes for signaling between distant organs in response to biotic or abiotic stresses, physiological adjustments and developmental switches. The phloem plays a major role in this long distance signaling pathway, by transporting a range of signal molecules, including hormones, metabolites and macromolecules (Van Bel, 2003; Dinant et al., 2010). The phloem sap translocation stream involves highly specialized cells forming the sieve elements, which consist of enucleated cells with no known transcriptional or translational activity when fully differentiated (Sjölund, 1997). Most of the RNA and protein molecules found in sieve elements are thought to be transported from the neighboring companion cells, through symplasmic connecting plasmodesmata (Lucas and Lee, 2004). Some of the proteins in the sieve element probably remained anchored to the parietal layers and help to maintain the cell machinery (Van Bel, 2003), but most macromolecules are transported by mass flow to distant organs, where they may have non cell-autonomous effects on gene expression (Banerjee et al., 2004; Lough and Lucas, 2006; Corbesier et al., 2007). This model assumes the existence of an efficient delivery pathway between companion cells and sieve elements (Kim, 2005) and a regulatory trafficking control system for the long-distance targeting of macromolecules. A destination-selective mechanism was also proposed for the trafficking of proteins in sieve elements (Aoki et al., 2005). A few non cell-autonomous proteins (NCAP) binding to phloem sap proteins and acting as chaperones have been implicated in this long-distance trafficking pathway (Aoki et al., 2002; Lee et al., 2003; Aoki et al., 2005; Taoka et al., 2007; Ham et al., 2009).

Despite the recent descriptions of large sets of macromolecules, including proteins, mRNAs and small RNAs, in phloem sap (Kehr, 2006; Kehr and Buhtz, 2008; Dinant and Lemoine, 2010), our understanding of the mechanisms of transport of these molecules into and through sieve elements remains incomplete. Sieve elements have a unique cellular organization, in which the plasma membrane and smooth parietal endoplasmic reticulum (ER) are retained but the other endomembrane systems, such as the Golgi apparatus, the tonoplast, and the nucleus, disappear during differentiation (Van Bel, 2003). There are also no organized microtubules and microfilaments (Evert, 1990). The trafficking of macromolecules therefore presumably
depends on other as yet unidentified networks, regulating the entrance and exit of macromolecules and preventing the obstruction of sieve pores.

P-proteins (for phloem proteins) are idiosyncratic sieve element structures thought to be involved in regulatory processes within sieve elements. They are defined as large proteinaceous bodies observed on transmission electron microscopy and found in the sieve elements of dicotyledonous and many monocotyledonous species (Evert, 1990). Depending on the species and stage of development, P-proteins may be observed as filaments, tubules or crystalline structures (Esau, 1969). It has been proposed that P-proteins are involved in defenses against pathogen colonization of the phloem or against sap-sucking insects, such as aphids. They are thought to occlude the sieve pores in case of injury (Spanner, 1978). Light and confocal laser scanning microscopy studies have demonstrated that forisomes — crystalloid P-proteins found in *Vicia faba* and other faboid legumes — control the conductivity of sieve tubes (Knoblauch *et al.*, 2001; Knoblauch *et al.*, 2003). The components of the forisomes have been identified and appear to be proteins unique to faboid legumes, although some of the motifs they contain have been identified in proteins from other plant taxa (Noll *et al.*, 2007; Pelissier *et al.*, 2008). These findings demonstrate a role of forisome-type P-proteins in the occlusion of sieve tubes in the Fabaceae, but the function of the P-proteins present in other species remains unclear.

The composition of filamentous P-proteins has been determined in cucurbits, from which large amounts of phloem sap can be collected. In *Cucurbita maxima*, the P-proteins are the two most abundant sap proteins: phloem protein 1 (PP1), which belongs to a gene family found only in cucurbits (Clark *et al.*, 1997), and phloem protein 2 (PP2), which belongs to a large gene family in angiosperms (Bostwick *et al.*, 1992; Bostwick *et al.*, 1994; Dinant *et al.*, 2003). In cucurbits, PP2s display additional properties, including RNA binding (Gómez and Pallás, 2001; Owens *et al.*, 2001; Gómez and Pallás, 2004; Gómez *et al.*, 2005), trafficking from cell to cell and modification of the size exclusion limit of the plasmodesmata (Balachandran *et al.*, 1997). These observations suggest that PP2 might play a role in the cell-to-cell trafficking of macromolecules. In *C. maxima*, PP2 is produced in companion cells, but the protein accumulates in sieve elements (Dannenhoffer *et al.*, 1997) and is transported across long distances (Golecki *et al.*, 1998; Golecki *et al.*, 1999). Early studies suggested that the P-proteins of *C. maxima* exist in at least two structural states: large polymers immobilized in individual sieve elements and small polymers or individual subunits translocated over long
distances (Read and Northcote, 1983a). A phloem-specific pattern of expression has been reported for the PP2 genes in two other plant species: *Apium graveolens* (AgPP2-1), a member of the Apiaceae, and *Arabidopsis thaliana* (AtPP2-A1 and AtPP2-A2), a crucifer. In both species, these genes were found to be expressed specifically in the companion cell-sieve element complex (Dinant et al., 2003). PP2 transcripts have also been found in phloem transcriptomes from various species (Ivashikina et al., 2003; Vilaine et al., 2003; Schrader et al., 2004; Zhao et al., 2005; as discussed in Le Hir et al., 2008) and more recently in the companion cell translatome in Arabidopsis (Mustroph et al., 2009).

The PP2s isolated from *C. maxima* and *Cucumis melo* have lectin activity (Sabnis and Hart, 1978; Allen, 1979; Read and Northcote, 1983b; Dinant et al., 2003). These proteins bind to chitin columns, demonstrating an affinity for oligomers of N-acetyl-glucosamine. These observations led Allen (1979) and Read and Northcote (1983) to suggest a role in defenses against fungi or bacteria. Other lectins that bind to oligomers of N-acetyl-glucosamine have been identified in plants, and interactions with glycoproteins from plant pathogens, including fungi, have been demonstrated for some of these molecules (Peumans and Van Damme, 1995). Plant lectins also frequently display insecticidal activity (Carlini and Grossi-de-Sa, 2002; Vasconcelos and Oliveira, 2004). Furthermore, it has recently been suggested that carbohydrate-binding properties are involved in endogenous signaling processes (Van Damme et al., 2004), although little is known about the role of PP2 in the recognition of endogenous or foreign glycosylated molecules.

Despite the large body of information available, studies on the properties and functions of PP2 phloem lectins have been limited thus far to cucurbits. This is mainly because of the phloem properties of cucurbits, which exude large volumes of phloem sap (Turgeon and Wolf, 2009) and have higher concentrations of proteins in phloem sap, up to 100 mg.mL⁻¹ (Richardson et al., 1982), than the plants of other families (0.1–2.0 mg.mL⁻¹) (Thompson and Schulz, 1999). However, the composition of P-proteins and functions of PP2 proteins may differ in other species. We therefore investigated the properties of an Arabidopsis phloem PP2, PP2-A1. We produced large amounts of PP2-A1 protein in *E. coli*, making it possible to analyze in detail the carbohydrate-binding properties of PP2-A1. We also investigated the binding of PP2-A1 to phloem sap proteins and its effect on aphids in an *in vitro* diet assay with the pea aphid *Acyrthosiphon pisum* (for which Arabidopsis is a non host species) and the green peach aphid *Myzus persicae* (for which Arabidopsis is a host species). The results of these experiments,
together with those for intrinsic fluorescence spectroscopy and surface plasmon resonance, indicated that PP2-A1 contains several carbohydrate-binding sites with potentially different functions in the trafficking of endogenous proteins or interactions with sap-sucking insects.

Results

Translational fusion of PP2-A1 with a His6 tag

The PP2-A1 gene (Genbank accession no. At4g19840.1) encodes a 246-amino acid protein (28.1 kD; accession no. O81865), as predicted from the identification of an open reading frame from the sequences of the PP2-A1 cDNA (Dinant et al., 2003) and of full-length cDNAs available from Genbank (AY085730; AY090355; BX829358; AY122906, BT015460). Arabidopsis thaliana produces only very small amounts of phloem sap by exudation (Barnes et al., 2004). We therefore generated recombinant PP2-A1 by heterologous expression in E. coli. This bacterium is a suitable host because there is no evidence that PP2 phloem lectins require glycosylation for activity in vivo (Allen, 1979; Balachandran et al., 1997). The PP2-A1 cDNA was amplified from Arabidopsis total RNA and the coding region was fused in frame with a 6–His tag (His6) in the C-terminal position, to facilitate its subsequent purification on a Ni2+-nitriloacetic affinity column. This sequence was introduced into the E. coli expression vector pET9, for expression in the Tuner DE3 pLys S strain.

Production of recombinant PP2-A1-His6

The recombinant PP2-A1-His6 was produced in E. coli grown in a large-scale fermentor. The protein was purified in a one-step process, on a nickel affinity column. Recombinant PP2-A1-His6 resolved as a doublet with an apparent molecular mass of about 28 kD on a denaturing gel (data not shown). Mass spectrometry (ESI-MS) data confirmed the presence of two polypeptides of 27.294 kD and 28.7895 kD (mass accuracy of 0.01%). These proteins had overlapping sequences, as confirmed by MALDI-MS analysis after treatment with trypsin, and together they covered up to 67 % of the PP2-A1 sequence. The two polypeptides differed only in terms of their N-terminal ends. The largest protein corresponded to the predicted PP2-A1 sequence (including a HCSELLPNK peptide corresponding to codons 5-12), whereas the shorter peptide lacked the first 13 N-terminal amino acids. This shorter product was the most abundant and its sequence began just after a second in-frame Met initiation codon, located
exactly 13 codons downstream from the predicted translation initiation codon (Supplementary Figure S1). This observation is consistent with the new revised annotation (Alexandrov et al., 2006), which has reassigned the start codons of a number of Arabidopsis genes, including PP2-A1, based on a more favorable consensus sequence for the initiation of translation (accession number: AAM62948). The shorter product was consistent with a theoretical mass of 27.2964 kD after posttranslational removal of the N-terminal Met in E. coli. The mass of the larger product (28.7895 kD) was also consistent with the removal of the N-terminal Met in E. coli (theoretical mass of 28.7932 kD) (Supplementary Figure S1).

A second construct, based on the second initiation codon, with the coding sequence in translational fusion with a C-terminal His6 tag, was generated to ensure that a homogeneous product was obtained. This second recombinant PP2-A1-His6 was produced according to the same procedure. After purification, a single product was obtained with a mass of 27.4282 kD, as determined by mass spectrometry, consistent with the theoretical molecular mass of 27.4276 kD. The two constructs yielded similar amounts of proteins (mean of 100 mg.L⁻¹) with similar mobility on electrophoresis in denaturing gels (data not shown). The shorter version of the recombinant PP2-A1-His6 was used in all the experiments described below, unless otherwise specified.

**PP2-A1-His6 binds to a chitin column**

A chitin column was used to determine the capacity of PP2-A1-His6 to interact with oligomers of GlcNAc. This affinity column consisted of GlcNAc polymers. The recombinant PP2-A1-His6 was produced in bacteria, and the crude bacterial extract was applied to a GlcNAc-Sepharose column. After elution, the various bound and unbound fractions were analyzed by SDS-PAGE (Figure 1). Most bacterial proteins were not retained on the column after stringent elution in alkaline conditions. The main protein recovered was a 27.5 kD product with an electrophoretic mobility similar to that of the recombinant PP2-A1-His6 purified on the nickel affinity column (Figure 1, lane 8). The recombinant protein therefore had a strong capacity for binding oligomers of GlcNAc. The bound fraction yielded a second band of about 56 kD. This product may correspond to a dimer.

**PP2-A1-His6 binds to GlcNAc oligomers, high-mannose N-glycans and a sialic acid**
We demonstrated specific binding of PP2-A1-His6, using the glycan array developed by the Consortium for Functional Glycomics (www.functionalglycomics.org). Version 3.1 of this array, consisting of 377 different glycans, was used to determine the specificity and relative affinity of PP2-A1-His6 for oligosaccharides (Blixt et al., 2004). In this assay, protein binding was quantified with a fluorescent antibody against the His6 tag present on the recombinant protein. PP2-A1-His6 was tested at several concentrations (0.1, 10, 50 and 100 µg.mL⁻¹). At a concentration of 0.1 µg.mL⁻¹, no specific interaction was observed. At concentrations of 10 and 50 µg.mL⁻¹, the signal increased linearly with protein concentration, indicating the relative affinity of PP2-A1-His6 for the various glycans. At a higher concentration (100 µg.mL⁻¹), strong binding to a large number of glycans was observed, indicating saturation.

The nine glycans for which fluorescence was strongest (largest number of RFUs) at a concentration of 50 µg.mL⁻¹ are displayed in Figure 2. They were classified into three classes. Short oligomers of N-acetylglucosamine (GlcNAc) were recognized; PP2-A1 preferentially bound N-acetyl (Ac) chitotriose (GlcAcβ1-4GlcNAcβ1-4GlcNAcβ) and displayed moderate binding to N-Ac chitopentose (GlcNAcβ1-4)₅ and N-Ac chitohexose (GlcNAcβ1-4)₆. Monomers and dimers of GlcNAc were not recognized. Several high-mannose N-glycans were also strongly recognized by PP2-A1-His6. These molecules were essentially sugars with a common minimal structure, Man₃α₁-3(Man₆α₁-6)Manβ₃-4GlcNAcβ₃-4GlcNAcβ (Man₃GlcNAc₂), associated with additional Man α3 or α6 linkages. PP2-A1-His6 also bound to the sialic chain 9-acyl-N-acetylneuraminic acid (9AcNeu₅Ac), although neither Neu₅Ac nor other sialic acids present on the glycan array were recognized. Comparison of the binding responses at different protein concentrations demonstrated that PP2-A1-His6 had the highest affinity for high-mannose N-glycans, oligomers of GlcNAc and 9AcNeu₅Ac (Figure 3). No other glycans, including complex-N-glycans or other O-linked glycans, were recognized (Supplementary Table S2). These findings provide evidence for the binding of PP2-A1-His6 to glycans with different structures.

Surface plasmon resonance analysis of binding properties
We studied the binding of PP2-A1-His<sub>6</sub> and N-N’-N”triacetylchitotriose by surface plasmon resonance experiments assessing the capacity of PP2-A1-His<sub>6</sub> to interact with oligomers of GlcNAc. N-N’-N”triacetylchitotriose is a commercially available structural analog of chitotriose that is ideal for structural studies. The sensorgrams obtained after PP2-A1-His<sub>6</sub> immobilization on the CM5 sensorchip, with the analyte, N-N’-N”triacetylchitotriose, circulating at different concentrations are shown in Figure 4A. Differences in response between the response channel, where interaction took place, and the reference response channel, where no interaction occurred, were assessed as a function of time. We found that PP2-A1-His<sub>6</sub> readily bound N-N’-N”triacetylchitotriose. The relative response, as a function of N-N’-N”triacetylchitotriose concentration and displayed as lin-lin and log-log representations (Figure 4B), reached a plateau at high sugar concentrations. Thus, all the binding sites were saturated for N-N’-N”triacetylchitotriose concentrations exceeding 600 μg.mL<sup>-1</sup>. The response was not linear and the curve did not have a sigmoid shape at higher sugar concentrations. These data are consistent with the involvement of several binding sites per molecule in the interaction of PP2-A1-His<sub>6</sub> with N-N’-N”triacetylchitotriose. We analyzed the stoichiometry of the interaction, by comparing the curve for the binding of N-N’-N”triacetylchitotriose to immobilized PP2-A1-His<sub>6</sub> with the theoretical maximum binding capacity of the immobilized ligand (R<sub>max</sub>). For a stoichiometry of 1:1, an R<sub>max</sub> close to 45 RU was predicted. For concentrations of N-N’-N”triacetylchitotriose greater than 400 μg.mL<sup>-1</sup>, we obtained values exceeding 45 RU, consistent with the binding of more than one molecule of N-N’-N”triacetylchitotriose per molecule of PP2-A1-His<sub>6</sub>. For a stoichiometry of 2:1, the predicted R<sub>max</sub> value was about 90 RU. The experimental values obtained were slightly lower, but were nonetheless best fitted by this model. These findings suggest that binding sites may not have been entirely accessible for interaction with N-N’-N”triacetylchitotriose, because of the random orientation of the proteins after binding via amine coupling to carboxymethylidextran.

**Assessment of binding site characteristics by intrinsic fluorescence spectroscopy**

Binding properties were further assessed by titration experiments, using intrinsic fluorescence spectroscopy. The binding of N-N’-N”-triacetyl-chitotriose to PP2-A1-His<sub>6</sub> reduced tryptophan fluorescence intensity without changing the maximum emission wavelength, λ<sub>max</sub>, which remained at 344 nm (Figure 5A). This quenching of tryptophan fluorescence probably
reveals a slight change in the conformation of PP2-A1-His$_6$ following interaction with sugar moieties. The absence of a change in maximum fluorescence emission wavelength ($\lambda_{\text{max}}$) suggests that tryptophan residues may not be fully exposed to aqueous solvent, as previously reported in other binding systems (Renard et al., 1998). A hyperbolic fluorescence titration curve was obtained when the results were plotted as a function of sugar:protein molar ratio (Figure 5B). The dissociation constant ($K_d$) and the number of sugar binding sites ($n$) were determined by fitting a Scatchard equation to the experimental curve, assuming a non cooperative model with independent and non identical sites (Figure 5 C). The $K_d$ (slope) and $n$ (intercept) values obtained from linear regression analysis suggested that N-N’-N”triacetylchitotriose bound PP2-A1-His$_6$ through two different types of site: a high-affinity site with a $K_d$ of $3.4 \times 10^{-7}$ M and an $n$ value of 0.5, and a low-affinity site with a $K_d$ of $1.3 \times 10^{-5}$ M and an $n$ value of 1.9. These values for the number of binding sites suggest that two proteins (i.e. a dimer) bind one sugar via the high affinity sites, whereas two sugars bind one protein via the low affinity sites. The existence of two classes of site on PP2-A1 suggested that the interaction of N-N’-N”triacetylchitotriose with PP2-A1 promoted the formation of a protein dimer ($n$=0.5), with additional sugar molecules loosely binding to other regions of PP2-A1 ($n$=1.9). Given the high affinity of the binding site promoting the dimer formation, PP2-A1-His$_6$ probably exists as as dimers or oligomers in solution.

**Prediction of the structure of the carbohydrate binding domain**

Searches of the Pfam $\alpha_2$– and Prosite motif servers detected no lectin motifs in PP2-A1 or other related PP2s (CbmPP2, CmmLec 17 and CmmLec 26). We therefore tried to identify structural similarities to proteins of known structure, by Blast searches, using PP2-A1 as the query and the CBS @tome2 metaserver against protein structure databases. PP2-A1 presented significant structural similarities (TITO scores < -60) to two classes of carbohydrate binding modules (CBM), the Cbm4 of the laminarase 16a (lam16a) from *Thermotoga maritima* and the Cbm22-2 of the endo-1,4-beta-xylanase Y (Xyn10B) from *Clostridium thermocellum*. These domains had a $\beta$-strand structure (Xie et al., 2001; Boraston et al., 2002) typical of carbohydrate binding domains. In Cbm4 and Cbm22-2, the carbohydrate binding site is located within a groove, in a structure consisting of parallel and anti-parallel $\beta$-strands. Alignments of the sequences of PP2-A1 and of these CBM domains identified an overlap in
the area corresponding to the structural β-strands. In PP2-A1, structural similarities were found in the sequence from residues 112 to 249, making up part of the conserved core of the protein, including part of domain A and domains B to D (Dinant et al., 2003). Based on the similarity to Cbm4, which gave the best consensus evaluation with the TITO program, a model for the structure of PP2-A1 was proposed (Figure 6). In this model, the β-strands were positioned on the conserved core and constituted a pocket-like structure consisting of parallel or anti-parallel β-strands. This region of similarity to CBM may be the putative lectin binding site. Unlike lam16a and Xyn10B, PP2-A1 had no recognizable calcium binding domain (data not shown). Structural predictions were not possible for the N-terminal region of PP2-A1 (residues 1-111).

**PP2-A1-His$_6$ binds to phloem sap proteins**

As PP2-A1 was potentially present in the phloem sap, we investigated the binding of PP2-A1-His$_6$ to phloem sap proteins. Phloem sap was obtained from mature rosette leaves with an EDTA-facilitated exudation technique (King and Zeevaart, 1974) adapted for Arabidopsis. The quality of the sap was confirmed by the smaller proportions of glucose and hexose than of sucrose (Figure 7A), as expected for phloem sap exudate. In addition, the phloem sap protein profile was clearly different from the profile for total proteins extracted from the leaf veins (Figure 7B), further indicating low levels of contamination from other tissues. On SDS-PAGE, major bands were observed at 27 and 55 kD. These fractions were isolated from excised bands and analyzed for the characterization of abundant sap proteins. The proteins were analyzed by LC MS/MS spectrometry. We unambiguously identified 24 proteins. The 27 kD fraction was significantly enriched in various isoforms of glutathione S-transferase and dehydroascorbate reductase (Supplementary Table S3). The 55 kD fraction was enriched in catalase, enolase phosphopyruvate hydratase, adenosylhomocysteinases, lipoamide dehydrogenase and several leucine-rich repeat (LRR) proteins. Three proteins matched phloem sap proteins previously identified in phloem sap from a related species, *Brassica napus* (Giavalisco et al., 2006): the glutathione S-transferase GSTF9, the adenosylhomocysteinase 1 HOG1 and the UDP-glucose pyrophosphorylase UGP2. In addition, we found peptides covering the elongation factor 2 LOS1 and the elongation factor alpha 1 MUF9.8 also reported in the phloem sap of *B. napus* (data not shown). As this analysis covered only the 27 and 55 kD fractions, it probably covered only a fraction of the
phloem sap proteome, overlapping only partially with the proteomes of other species, such as *B. napus*. PP2-A1 was not found in these fractions, although 27 and 55 kD are close to the theoretical masses for PP2-A1 monomers and dimers, respectively.

Far-western analysis was performed to determine whether PP2-A1-His\(_6\) interacted with phloem sap proteins. PP2-A1-His\(_6\) bound several sap proteins (*Figure 7C, lane 4*). At least twelve bands were revealed by protein overlay with PP2-A1-His\(_6\). The most intense bands corresponded to proteins with apparent molecular masses of approximately 24-26, 43 and 55-60 kD. After glycosidase treatment to remove N-linked glycans (PNGase) or tGlcNAc (GDase) (*Figure 7C*, lanes 5 and 6, respectively), PP2-A1-His\(_6\) still bound to phloem sap proteins, although closer examination revealed slight modifications to the binding pattern. PNGase treatment decreased the intensity of labeling of bands in the 24-26 kD and 55-60 kD fractions. Some bands in the 55-60 kD fraction were less strongly labeled after GDase treatment. Thus, although glycosidase treatments slightly altered the binding spectra of PP2-A1-His\(_6\), they did not abolish its binding to phloem sap proteins.

**Effect of PP2-A1-His\(_6\) on the growth of aphids**

The effect of PP2-A1-His\(_6\) on insect weight gain and survival was assessed in an insect feeding bioassay. This experiment was conducted either with the pea aphid *A. pisum*, used as a reference aphid for lectin toxicity tests (Rahbé and Febvay, 1993), or with the green peach aphid *M. persicae*, an aphid infesting Arabidopsis. Aphid nymphal weight and mortality were assessed after the neonates fed for seven days on an artificial diet supplemented with various concentrations of recombinant PP2-A1-His\(_6\). Wheat germ agglutinin (WGA), a GlcNac lectin of limited toxicity to aphids (Rahbé and Febvay, 1993), and casein, a protein with no aphid toxicity, were used as controls in this experiment.

After seven days, no significant difference in survival was observed between *A. pisum* and *M. persicae* nymphs fed on the control diet and nymphs fed on the diet supplemented with recombinant protein, regardless of the concentration of recombinant protein added to the medium (*Supplementary Table S4*). However, aphid weight was reduced at protein concentrations between 125 and 250 µg.mL\(^{-1}\) for *A. pisum* (P < 0.001 and P = 0.0005, for 125 and 250 µg.mL\(^{-1}\), respectively) (*Figure 8A*) and between 83-250 µg.mL\(^{-1}\) for *M. persicae* ,
with the L06 clone (P < 0.003 for first biological replicate 1 and P=0.01 for second replicate) (Figure 8B). Similar results were obtained on a second clone of *M. persicae*, the LC05 clone (Supplementary Figure S6). The loss of weight was more pronounced in *A. pisum* (33% maximal weight decrease) than in *M. persicae* (10 to 20% maximum weight decrease). At higher concentrations (500-1000 µg.mL⁻¹), weight gain was unaffected in both *A. pisum* and *M. persicae*. No effect on the weight and survival of either aphid species was observed with WGA and casein (Figure 8C).

In this bioassay, the recombinant protein had no effect on nymphaI settlement in *A. pisum* or *M. persicae*, which was similar after 24 h on artificial diets with or without recombinant protein supplementation, for all doses (Supplementary Table S4). This demonstrates that the recombinant protein had no deterrent effect in a no choice assay. Similarly, in a choice assay in which the recombinant protein was present at a concentration of 200 µg.mL⁻¹, there was no phagodeterrent effect after 16 h, either in *A. pisum* or in *M. persicae* (Supplementary Table S5), showing that the effect on weight gain did not result in changes in insect behavior. In such a choice test with the artificial diet supplemented with 500 µg.mL⁻¹ of recombinant protein, we observed a significant “deterrent” effect after 16 h with *A. pisum* (p=0.0004), characterized by a negative phagostimulation index (Ix=-0.53; i.e. ≈ 23% aphids feed on the medium supplemented with PP2-A1-His6). However, no effect was observed with either of the two clones of *M. persicae* used (Supplementary Table S5).

**Discussion**

The phloem constitutes a strategic checkpoint for the mounting of defenses against pathogens and pests (van Bel and Gaupels, 2004). It provides an efficient pathway for long-distance communication and the trafficking of macromolecules, although our knowledge of the phloem proteins involved in these processes remains incomplete (Lough and Lucas, 2006). PP2s constitute an interesting family of phloem proteins. They were first identified as components of the so-called “P-proteins” in cucurbits and shown to act as dimeric chitin-binding lectins in these plants (Beyenbach *et al.*, 1974; Sabnis and Hart, 1978; Allen, 1979; Read and Northcote, 1983a; Smith *et al.*, 1987; Bostwick *et al.*, 1992), leading to the suggestion of a role in defense mechanisms. In *C. maxima*, CbmPP2 presents several properties associated with NCAPs (Bostwick *et al.*, 1994; Balachandran *et al.*, 1997; Golecki *et al.*, 1998; Golecki
et al., 1999), suggesting a role in the trafficking of macromolecules. However, it remains unclear whether PP2s from species other than cucurbits present such properties. We demonstrated that lectin activity is conserved in Arabidopsis, for the closest ortholog, PP2-A1. We then extended these results by characterizing the spectrum of sugar recognition, the properties of the binding sites and by demonstrating binding to several phloem sap proteins. In addition, we propose a model for the structure of PP2-A1, based on structural predictions.

Structure prediction
The PP2 family is unique to plants and is unrelated to other classes of plant lectins. Structural database searches revealed similarity to two carbohydrate-binding modules (CBM) comprising the catalytic domain of two bacterial enzymes involved in glycan modifications. The PP2-A1 region that aligned with the CBM pocket for sugar binding (Figure 6) is predicted to fold into a parallel and anti-parallel β-strand structure. This region overlaps the conserved region (residues 112-246), which includes part of the domain A and the domains B to D (Dinant et al., 2003). It is unknown whether this predicted carbohydrate binding module corresponds to the domain required for the lectin activity associated with PP2-A1. However, all the lectin domains crystallized to date have a structure based on β-strands, supporting the hypothesis that the domain predicted to fold into β-strands in PP2-A1 may be responsible for its lectin activity. It was not possible to assign a structure to the first 111 amino-terminal residues of the protein. This N-terminal region contains no known motif. However, other PP2 proteins have an extension at their N-termini, often containing protein-protein interaction motifs, such as the F-Box, TOLL domain or AIG1 domain (Dinant et al., 2003). Thus, despite the absence of any known domain in PP2-A1, this region may also contain a binding domain that has yet to be identified.

PP2-A1 is a GlcNAc oligomer-binding lectin with two types of binding site
The chitin-binding assay showed that PP2-A1-His6 produced in E. coli bound GlcNAc oligomers (Figure 1), as reported for the cucurbit PP2 phloem lectins. Glycan array screening further extended these results by showing that PP2-A1-His6 preferentially recognized (GlcNAc)₃ chitotriose and had a lower affinity for (GlcNAc)₅ and (GlcNAc)₆ oligomers. No recognition was observed for GlcNAc dimers and monomers (Figure 2). The specificity of the interaction with (GlcNAc)₃ chitotriose was confirmed by surface plasmon resonance analysis (Figure 4) and characterized further by intrinsic fluorescence spectroscopy based on titration with N,N’,N”triacetylchitotriose (Figure 5). We identified two classes of binding
site. Each PP2-A1 dimer had one high-affinity binding site ($K_d = 3.4 \times 10^{-7}$ M, $n = 0.5$) and each monomer had two low-affinity sites ($K_d = 1.3 \times 10^{-5}$ M, $n = 1.89$). This result is important, as it demonstrates that sugar binding is closely connected to dimer formation.

PP2 lectins from cucurbits and the related protein nictaba have been reported to be dimeric in planta (Read and Northcote, 1983a; Chen et al., 2002; Walz et al., 2004). The existence of a high-affinity site in PP2-A1 dimeric forms is consistent with these observations. However, it remains unclear whether dimer formation is a prerequisite for the binding of sugar moieties or whether sugars are required for the formation of dimeric PP2-A1. The binding of PP2-A1 to other sugars (see below) probably renders the system even more complex, because these sugars may also have different affinities for PP2-A1 subunits and dimers. This raises the question of the conditions required for oligomerization in vivo and the role of PP2 subunits and aggregates in planta, as previously considered for the cucurbit PP1 and PP2 phloem proteins (Read and Northcote, 1983a; Leineweber et al., 2000).

**PP2-A1 binds to different classes of glycans**

Glycan array screening revealed that, in addition to GlcNAc oligomers, PP2-A1-His$_6$ bound to high-mannose N-linked glycans and to a sialic acid derivative (Figure 2). This extends the specificity of the phloem PP2-A1 lectin to previously unsuspected classes of glycans. PP2-A1-His$_6$ bound 9NAcNeu$_5$Ac with high affinity. Sialic acids are present neither in plants nor in their pathogens, although a few sialic acid-specific plant lectins are known (Lehmann et al., 2006; Zeleny et al., 2006). Sialylation occurs mostly in mammals, birds and their pathogenic microorganisms, in which it acts in recognition phenomena (Vimr and Lichtensteiger, 2002; Schauer, 2009). The recognition of 9NAcNeu$_5$Ac by PP2-A1 may thus facilitate binding to sialylated proteins from animal herbivores, as suggested for other plant lectins from this group (Lehmann et al., 2006).

PP2-A1-His$_6$ also bound high-mannose N-linked glycans, with a common Man$_3$GlcNAc$_2$ core structure, which was recognized with high affinity (Figure 3). High-mannose N-glycans are found in animals, plants, plant pathogens and insects. High-mannose N-glycans are processed in the ER (Lerouge et al., 1998). In plant cells, oligomannoside glycans account for only a small fraction of N-glycans and Man$_3$GlcNAc$_2$ is rare (Strasser et al., 2004; Strasser et al., 2005). The secretory pathway of sieve elements is limited (Van Bel, 2003), potentially
limiting the formation of complex glycans and leading to an enrichment in glycoproteins with high-mannose N-glycans. It is therefore possible that PP2-A1-His6 recognizes phloem sap glycoprotein. Man3GlcNAc2 is also the major processed N-glycan produced by insect cells (Hollister et al., 2002; Tomiya et al., 2004); insects are thus an obvious potential target (see below).

Based on our glycan array analysis, it now appears that PP2-A1-His6 binds different classes of glycans. This result raises the interesting possibility that the phloem PP2 lectin plays several roles, in recognizing either endogenous glycoproteins or glycosylated receptors of pathogens and herbivores feeding on phloem sap.

**A role for PP2-A1 in the recognition of phloem sap t-GlcNAc proteins?**
The binding of PP2-A1-His6 to small oligomers of GlcNAc is consistent with previous reports for other PP2 lectins (Sabnis and Hart, 1978; Allen, 1979). PP2-A1-His6 did not bind to long GlcNAc oligomers and is therefore unlikely to be involved in recognizing the chitin found in insects and fungi. Instead, it may be involved in the recognition of O-linked GlcNAc, such glycans being found in glycoproteins from many organisms. Terminal GlcNAc are found as monomers in animal cells (Hanover, 2001; Wells and Hart, 2003) and as oligomers in plant cells (Heese-Peck et al., 1995). GlcNAc glycosylation occurs in the cytoplasm and is involved in various signaling pathways and nucleocytoplasmic cycling (Hart et al., 2007). Phloem sap and plant viral proteins can be modified by O-GlcNAc glycosylation (Chen et al., 2005; de Jesus Perez et al., 2006; Scott et al., 2006; Taoka et al., 2007). Non cell-autonomous pathway protein 1 (Nt-NCAPP1), which is required for the trafficking of glycosylated proteins through plasmodesmata (Taoka et al., 2007), binds O-GlcNAc proteins in phloem sap. The phloem Hsp70 is another candidate protein for binding to O-GlcNAc phloem sap proteins (Aoki et al., 2002). These studies led the authors to propose that transport, through plasmodesmata, between companion cells and sieve elements, like shuttling between the nucleus and the cytoplasm, involves O-GlcNAc glycosylation.

Based on our current findings we suggest that PP2-A1 may bind O-GlcNac proteins. We established that PP2-A1-His6 bound to phloem sap proteins, based on the observation that PP2-A1-His6 interacted with several proteins, ranging from 20 to 80 kD in size (Figure 7). Moreover, our results indicate that the range of PP2-A1-His6 protein binding is broader than previously found, for example in *C. maxima*, in which PP2 was shown to interact only with
PP1 (Read and Northcote, 1983a; Read and Northcote, 1983b). The binding of PP2-A1-His$_6$ to phloem sap proteins was not entirely abolished by treatment with GDase (removing GlcNAc glycans) or PNGase (removing N-glycans), demonstrating that recognition does not strictly depend upon glycosylation. Such interactions may be involved in a range of processes, such as interaction with phloem sap proteins for their entry into the sieve elements, long-distance trafficking, signal transduction or folding. The recent observation that PP2-A1 enhances the transmission of a plant virus by aphids and stabilizes virus particles in vitro (Bencharki et al., 2010) is consistent with a role as a chaperone. PP2 is abundant in the phloem and is transported from companion cells (CC) to sieve elements (SE). It would therefore be abundant enough to chaperone the trafficking of glycoproteins from the CC to the SE, although additional investigations in vivo are required to confirm a number of hypotheses.

A role for PP2-A1 in the plant-aphid interaction?

Plant lectins are often insecticidal, as illustrated by the mannose-binding lectins, which act by binding to the glycoproteins of midgut epithelial cells (Vasconcelos and Oliveira, 2004). As the PP2-A1-His$_6$ protein bound to three classes of glycans, including Man$_3$GlcNAc$_2$, a major glycan in insect cells, we investigated the insecticidal effects of PP2-A1 on aphid feeding. Two species were tested: A. pisum, which is a model species for lectin toxicity (Rahbé and Febvay, 1993), and M. persicae, for which Arabidopsis is a common host. For this assay, we used a range of protein concentrations, from 16 to 1000 µg.mL$^{-1}$. Nothing is known about the concentration of phloem sap proteins in Arabidopsis and most other species. In rice, the concentrations of abundant proteins have been estimated at about 10 µg.mL$^{-1}$ (Suzui et al., 2006) and those of total soluble proteins have been estimated at about 150-200 µg.mL$^{-1}$ (Hayashi et al., 2000). Local concentrations of PP2-A1 in phloem sap are probably not homogeneous, because of local self-assembly. It therefore remains unclear whether the range of concentrations used in the artificial aphid bioassay, which are standard for toxicity tests, overlaps with local concentrations of PP2 in phloem sap.

When added to an artificial aphid diet, at concentrations of 125 to 250 µg.mL$^{-1}$ depending on the assay, PP2-A1-His$_6$ had a slightly detrimental effect on the weight gain of A. pisum nymphs (Figure 8A), which were up to 30% lighter than the controls after seven days of feeding. A similar effect was observed on M. persicae nymphs, with 83 to 250 µg.mL$^{-1}$ recombinant protein, although more moderate (nymphs 10 to 20% lighter) (Figure 8B). In this no choice test, aphid settlement and survival were not affected, demonstrating that PP2-
A1-His\textsubscript{6} was not acutely toxic at this range of concentrations (Supplementary Table S4). Similarly, at a concentration of 200 µg.mL\textsuperscript{-1}, no deterrent effect on any aphid species was detected in a choice assay (Supplementary Table S5). Thus, the slight growth-inhibitory effect was probably not due to a pre-ingestion effect. The smaller weight gain may instead result from the modulation of ingestion, with the presence of PP2-A1-His\textsubscript{6} in food preventing aphids from feeding abundantly. Alternatively, PP2-A1-His\textsubscript{6} may limit the aphid’s access to some essential nutrients, although there is no evidence that PP2s can bind nutrients present in phloem sap or in artificial diets for aphids, the composition of which is completely determined (holidic diet). Finally, this effect may result from the binding of PP2-A1-His\textsubscript{6} to glycoproteins in the midgut cells of the insects, or other internal targets, interfering with digestion, as observed for other plant lectins, potentially resulting in a post-ingestion antinutritional effect.

PP2-A1-His\textsubscript{6} had more moderate effects on M. persicae than on A. pisum, although each species reacted to the same range of protein concentrations. This suggests that adapted aphids are less susceptible to the effect of the lectin. Our results also confirmed the absence of an effect of the GlcNAc WGA lectin on aphid weight and survival (Figure 8C), as previously reported (Rahbé and Febvay, 1993). The observation that PP2-A1-His\textsubscript{6}, although not toxic for aphids, altered weight gain, suggests a mechanism different from those reported for GlcNAc lectins, such as WGA, which are harmless to aphids. It may instead result from its Man\textsubscript{3}GlcNAc\textsubscript{2} lectin activity, such moieties being abundant in insect glycoproteins. No significant effect was observed on weight gain for concentrations exceeding 500 µg.mL\textsuperscript{-1}, for the nymphs of both A. pisum and M. persicae. This lack of effect may result from aggregation of the recombinant protein in the artificial diet at such concentrations, making the protein less available for absorption from the diet. Protein aggregation has been observed in vitro at a concentration of 1000 µg/mL\textsuperscript{-1} in various buffers (unpublished). The observation that, at a recombinant protein concentration of 500 µg.mL\textsuperscript{-1}, A. pisum rejected the medium supplemented with PP2-A1-His\textsubscript{6} in a choice test also suggests that, at high concentrations, this protein renders the medium unattractive to feeding aphids. Protein solutions are unlikely to affect the gustatory perception of the diet by aphids (Sauvion et al., 2004), so the simplest explanation would be that PP2-A1-His\textsubscript{6} alters the viscodynamic properties of the diet.

Further investigations to elucidate the role of PP2-A1 on aphid growth would benefit from the use of transgenic plants overexpressing the PP2-A1 gene or pp2-a1 knockout mutants. One pp2-a1 insertion line has been described (Divol et al., 2007). Unfortunately, the level of
expression of the PP2-A1 gene was not affected by the insertion, which was located upstream from the promoter region, and, consistent with this finding, no effect on aphid settlement was observed (Divol et al., 2007), making it impossible to draw any firm conclusions about the role of PP2-A1 in vivo.

**Survey of phloem sap proteins**

We checked the quality of the phloem sap isolated by EDTA-facilitated exudation, by carrying out a survey of high-abundance phloem sap proteins based on LC MS/MS mass spectrometry. From the 27 and 55 kD fractions, we were able to characterize 24 proteins unambiguously (Supplementary Table S3). The relatively high frequency of glutathione S-transferase, dehydroascorbate reductase, catalase, copper ion binding oxidoreductase and enolase was consistent with the identification of oxidation-related or redox-related proteins in other reports of phloem sap protein analysis in other species (Walz et al., 2004; Giavalisco et al., 2006; Kehr, 2006). Three of the 24 proteins identified matched with orthologs found in the sap of *B. napus*: the UDP-glucose pyrophosphorylase UGP2, the glutathione S-transferase GSTF9 and the adenosylhomocysteinase SAHH1/HOG1 (Giavalisco et al., 2006). Several putative leucine-rich repeat (LRR)-related proteins were also found. Two LRR proteins were also found in pumpkin phloem sap (Lin et al., 2009). LRR domains are present in proteins with diverse functions, providing a framework for the formation of protein-protein interactions. In plants, they have been implicated in innate immunity signaling pathways (DeYoung and Innes, 2006). Several aphid resistance genes, conferring phloem-specific resistance, such the mi gene in tomato and the vat gene in melon, encode NBS-LRR proteins (Kaloshian et al., 2000; Dogimont et al., 2008), suggesting a potential role, in sieve elements, in interactions with phloem-feeding insects.

PP2-A1 was not detected in the 27 and 55 kD fractions, although its theoretical molecular mass would be consistent with detection in these fractions. A PP2 lectin was previously detected in *B. napus* phloem sap, with an observed molecular mass smaller than expected (Giavalisco et al., 2006). The absence of PP2-A1 in these fractions may have resulted from a mobility shift on SDS-PAGE due to posttranslational modifications. Alternatively, PP2-A1 may not be mobile in sieve elements, instead remaining bound to structural elements (Oparka and Cruz, 2000). Indeed, in pumpkin, it has been suggested that the phloem lectin attaches the P-proteins to the SE reticulum or plasma membrane (Smith et al., 1987), although there is also evidence to suggest that a fraction is translocated over long distances, through graft
unions (Golecki et al., 1998; Golecki et al., 1999). We also cannot rule out the possibility that the method used to collect phloem sap — involving the use of EDTA, a chelator of Ca²⁺ ions — affects the aggregation and mobility of PP2 proteins. Previous studies identifying PP2 phloem lectins in phloem sap proteomes were based on the collection of sap by phloem bleeding rather than EDTA-facilitated exudation (Walz et al., 2002; Walz et al., 2004; Giavalisco et al., 2006; Lin et al., 2009). It therefore remains unclear whether PP2-A1 subunits or small oligomers are mobile in the sieve elements of Arabidopsis, or whether they are attached to sieve element membranes or to P-proteins. The extent to which PP2 aggregates may depend on the cellular environment and may be modified following the entry of the protein into the sieve elements. Based on our findings that this protein has two classes of sugar binding sites, it will be interesting to investigate the subcellular distribution of this protein in vivo and to study the aggregation of PP2-A1 in different cell compartments or environmental conditions.

Materials and methods

Cloning of PP2-A1-His₆

Arabidopsis thaliana ecotype Columbia was used for the cloning of the PP2-A1 cDNA. Total RNA was isolated from frozen tissues, as previously described (Verwoerd et al., 1989). For initial cloning of the PP2-A1 cDNA, the coding sequence of the PP2-A1 cDNA previously amplified by RT-PCR from total RNA (Primers: forward A1-Fw: 5’ AAA AAA GCA GGC TCA TGA GCA AGA AAC ATT GCT CAG 3’; reverse A1-Rv 5’ AAG AAA GCT GGG TAC TGT TTG GGA CGA ATT GCA AC 3’) was reamplified by PCR, with forward primer A1NdeI (5’GGA ATT CCA TAT GAG CAA GAA ACA TTG CTC A 3’) and reverse primer A1NotI (5’TAA AGC GGC CGC AAA TCA GTG GTG GTG GTG CTG CTG TTT GGG ACG AAT TGC AAC 3’). These primers were ordered from Biomers (Ulm, Germany). They introduced an NdeI site at the 5’ end of the PCR product and a NotI restriction site at the 3’ end of the PCR product for further cloning (underlined sequences). PCR was performed in a GeneAmp 2700 PCR system from Applied Biosystems, with Taq DNA polymerase (New England Biolabs). The following PCR conditions were used: 3 min of denaturation at 94°C, 25 cycles of 45 s at 94°C, 30 s at 55°C and 1 min at 72°C, ending with an additional seven-minute period of elongation at 72°C. The PCR product was purified with the QIAquick Gel
Extraction kit (Qiagen) and inserted, as a NdeI/NotI fragment, into a modified pET9 vector (Novagen) into which a new NotI site had been introduced, to create the expression plasmid pET9-PP2-A1, for the expression of a translational fusion of PP2-A1 with a C-terminal six-histidine tag, under the control of the T7 promoter. For the shorter sequence of PP2-A1, forward primer A1Fw (5’ TTT AGA AAC CAA GAC TCG AAA 3’) and reverse primer A1bisNotI (TAA AGC GGC CGC AAA TCA GTG) were used to amplify the sequence from the pET9-PP2-A1 clone, which already contains the histidine tag sequence. The A1bis NdeI primer (GGA ATT CCA TAT GTT TAG AAA CCA AGA CTC GAA A) was used with the A1bis NotI primer, to introduce restriction sites for subsequent insertion into the pET9 vector and yielding pET9-PP2-A1s. PCR was performed as described above, with the following modifications: 30 cycles, with annealing at 54°C.

The plasmids were transferred into E. coli DH5α, in a transformation procedure involving the incubation of 150 ng of plasmid DNA with thermocompetent cells (Invitrogen) for 45 s at 42°C, followed by the addition of 2YT medium (4g.L⁻¹ bactotryptone, 2.5 g.L⁻¹ yeast extract and 1.25 g.L⁻¹NaCl), and incubation for 1 hour at 37°C. The cells were then resuspended in 2YT supplemented with kanamycin (50 µg.mL⁻¹) and incubated overnight at 37°C. The inserts present in the pET9-PP2-A1 and pET9-PP2-A1s plasmids were confirmed by sequencing (Millegen Service Facility, Labège, France), using 5’T7 promoter and terminator primers.

**Large-scale production of PP2-A1-His₆**

For large-scale protein production, the plasmids were introduced into E. coli Tuner DE3 pLysS cells (Novagen, Madison, USA). We used 10 mL of a fresh culture in 2YT medium supplemented with kanamycin (50 µg.mL⁻¹) and chloramphenicol (34 µg.mL⁻¹) incubated overnight at 37°C to inoculate 4 L of 2YT medium in a fermentor (Applicon) equipped for the maintenance of constant aeration levels, pH (7) and temperature (37°C). Production of the recombinant protein was induced by adding 1 mM isopropyl β-D thiogalactopyranoside (IPTG) to the culture when it reached an OD₆₀₀ of 1.5. After three hours of induction, the cells were harvested by centrifugation (20 min, 5000 x g) at 4°C. The pellet was frozen at −20°C. For purification purposes, we typically dissolved 5 mg of pellet (1/5 of the total pellet) in 30 mL of extraction buffer (50 mM Tris, 200 mM NaCl, 10 mM imidazole, pH 8). Two freeze-thaw cycles were carried out at -20°C/+20°C and -176°C (liquid nitrogen)+20°C, and the samples were sonicated for 90 s, with 10 s pulses of 80 W, at 4°C. The lysate was centrifuged...
for 45 min at 12000 x g and room temperature. The supernatant was then loaded onto a 30 mL fast-flow nickel affinity column (GE, Amersham) equilibrated with extraction buffer (50 mM Tris, 200 mM NaCl, 10 mM imidazole, pH 8). Proteins were eluted from the column (AKTA-PRIME, GE, Amersham) in an imidazole gradient of 250 to 500 mM in extraction buffer. Ten 5 mL fractions, corresponding to the protein elution profile, were typically collected after purification.

Fractions were dialyzed (Spectra/Por Membrane, MWCO: 6-8000) four times against 500 volumes of water, to remove salts and imidazole. The various fractions were then immediately frozen at -20°C and freeze-dried. Lyophilized PP2-A1-His₆ protein was stored in a desiccator in the dark. Proteins were analyzed by SDS-PAGE (Laemmli, 1970) in 12 or 15% Tris-HCl polyacrylamide gels. Before electrophoresis, the samples were boiled for 2 min in 2 x Laemmli sample buffer. After electrophoresis, the proteins were detected by staining the gels for 30 min at room temperature with Coomassie Brilliant Blue R (Bio-Rad).

Characterization of the recombinant protein by mass spectrometry
Mass spectrometry analyses were conducted at the “Biopolymers-Interactions-Structural Biology” platform of the INRA Center at Nantes (http://www.nantes.inra.fr/plateformes_et_plateaux_techniques/plateforme_bibs). The molecular mass of purified PP2-A1-His₆ was determined with an ion trap mass spectrometer (LCQ Advantage, Thermo-Fisher), equipped with an electrospray ionization source (ESI) operating in the positive ion mode. Samples were solubilized at a concentration of 20 nanomole.mL⁻¹ in an aqueous mixture of water and acetonitrile (1:1, v/v) acidified with 0.1% formic acid. Samples were infused into the mass spectrometer at a continuous flow rate of 5 μL.min⁻¹. Mass data were acquired for the mass-to-charge (M/z) range 400-2000.

The sequence of PP2-A1-His₆ was determined, after trypsin hydrolysis of the protein, by MALDI mass measurement for the resulting peptides (MALDI: matrix-assisted laser desorption ionization). We solubilized 50 µg of purified PP2-A1-His₆ in a 25 mM ammonium bicarbonate buffer (pH 8.5) and subjected the resulting solution to trypsin treatment (PROMEGA) for 3 hours at 37°C (enzyme: substrate ratio 1:30, w/w). The resulting mixture of peptides was diluted 1:50 in water. We mixed 1 µL of the dilution with 1 µL of the matrix preparation (2.5 mg.mL⁻¹ cyano-4-hydroxycinnamic acid, 2.5 mg.mL⁻¹ 2,5-dihydroxy benzoic acid, 70% (v/v) acetonitrile, and 0.1% (v/v) trifluoroacetic acid) and deposited the mixture on
the MALDI sample probe. MALDI-TOF mass spectrometry was performed for the 850-3500 \( M/z \) range, on a M@LDI LR instrument (Waters) equipped with a conventional nitrogen laser (337 nm) operating at a frequency of 10 Hz.

The smallest peptides generated by trypsin digestion were further analyzed by MS/MS fragmentation, based on nano-flow electrospray ionization tandem mass spectrometry (Nano-ESI-MS/MS) on a hybrid quadrupole orthogonal acceleration time-of-flight mass spectrometer (Q-TOF Global, Waters). The products of the trypsin digestion were diluted 1:100 in an aqueous mixture of water and acetonitrile (1:1, v/v) acidified with 0.1% formic acid. We introduced 2 \( \mu \)L of the resulting mixture into a metal-coated glass capillary (PROXEON) and allowed it to infuse into the mass spectrometer at an approximate flow rate of 100 nL.min\(^{-1}\). Mass data were acquired for the 300-2000 \( M/z \) range. The peptide at \( M_{H+} \) 1.0404 kD (doubly charged ion at \( M/z \) 520.7) was fragmented by collision-induced dissociation (CID) and fragments were measured over the 20-1200 \( M/z \) range. The fragment ions were compared with the theoretical b and y ions generated by the 5-13 sequence of PP2-A1-His\(_6\) (HCSELLPNK), using Mass Lynx Biotools software (Waters).

**Chitin column binding assay**

For this assay, we used the bacterial strains harboring the longer PP1-A1-His\(_6\) initially generated. The pellet recovered from 5 mL of culture, resulting in production of the recombinant PP2-A1-His\(_6\) protein, was solubilized in extraction buffer (50 mM Tris, 200 mM NaCl, pH 8) and cells were sonicated (1.5 min; 10 s pulse; 80 kHz; 4°C). The bacterial extract was applied to a 5 mL GlcNAc-Sepharose column (New England Biolabs). Unbound proteins were removed by washing the column twice with 5 mL of extraction buffer. Finally, successive washes with 5 mL of NaOH solutions at concentrations of 0.05, 0.1 and 0.5 M, resulted in desorption of the bound PP2-A1-His\(_6\). The eluted proteins were analyzed by SDS-PAGE in a 15% polyacrylamide gel, as described above, with detection by Coomassie Brilliant Blue staining.

**Glycan array binding affinity of PP2-A1-His\(_6\)**

PP2-A1-His\(_6\) was used to probe the printed glycan microarrays according to the standard procedure of Core H of the Consortium for Functional Glycomics (see https://www.functionalglycomics.org/static/consortium/resources/resourcemerecoreh8.shtml), Version 3.1 being used for the analyses reported here. This printed array consists of 377
glycans in replicates of six. Lyophilized recombinant protein was dissolved in Tris-buffered saline (20 mM Tris-HCl, 150 mM NaCl, 2 mM CaCl₂, 2 mM MgCl₂, pH 7.4 containing 0.05% Tween 20 and 1% BSA) to a concentration of 1 mg.mL⁻¹ and this solution was diluted to give a final concentration of 0.1, 10, 50 or 100 µg.mL⁻¹. Aliquots (70 µL) of PP2-A1-His₆ at the appropriate concentration were applied to each well of the printed microarray. The binding of PP2-A1-His₆ to specific glycans was detected and quantified as previously described (Subramanyam et al., 2008), except that recombinant PP2-A1-His₆ was detected with a commercial antibody against the His₆ tag. Binding was quantified by fluorescence techniques and the data are expressed as mean numbers of relative fluorescence units (RFU) for the replicates for each glycan represented on the array.

Surface plasmon resonance
Surface plasmon resonance (SPR) analyses (Achilleos et al., 2009) were performed on a BIACore 3000 optical biosensor equipped with a research-grade carboxymethyl dextran CM5 sensor chip (GE Healthcare). For PP2-A1-His₆ immobilization, the carboxyl groups of the CM5 sensorchip were activated at a flow rate of 20 µL.min⁻¹ with a 10-minute pulse of N-hydroxysuccinimide (NHS) and 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) (1:1 mixture). We then injected 100 µg.mL⁻¹ PP2-A1 in 10 mM acetate buffer, pH 5 for approximately 10 minutes and blocked the remaining succinimidyl esters by injecting 1 M ethanolamine hydrochloride, pH 8.5, for 10 minutes. The immobilization of about 2000 resonance units (RU) was achieved with the BIACore 3000 in these immobilization conditions. For binding experiments, the N-N’-N”triacetylchitotriose solution (Sigma) was diluted in 10 mM HEPES, 3 mM EDTA, 0.15 M sodium chloride, 0.005% P₂₀, pH 7.4 (HBS-EP) running buffer at various concentrations (0, 5, 10, 20, 50, 100, 150, 200, 400, 600, 800 and 1000 µg.mL⁻¹). The resulting solutions were passed over the PP2-A1-His₆ immobilized on the CM5 surface, at a flow rate of 20 µL.min⁻¹ and at a temperature of 25°C. Surface regeneration was achieved by injecting 50 mM NaOH into the system. PP2-A1-His₆ solution was injected over two channels, at a flow rate of 20 µL.min⁻¹. In one of these channels, the CM5 surface was activated with NHS-EDC (channel response), whereas, in the other channel, the surface was not activated (reference channel response). The specific response (or response difference) was obtained by subtracting the reference channel response from the channel response. Rₘₐₓ was defined as follows:

\[ R_{\text{max}} = \left( \frac{M_w \ N-N'-N''\text{triacetylchitotriose}}{M_w \ \text{PP2-A1-His}_6} \right) \times \text{RU immobilized} \times S \]
where $S$ is the stoichiometry. With N-N’-N”triacylchitotriose ($M_w = 627$) and PP2-A1-His$_6$ ($M_w \sim 28000$), $R_{max}$ is predicted to be close to 45 RU for a stoichiometry of 1:1 \[\left(\frac{627}{28000}\right) \times 2000 \times 1 = 44.8 \text{ RU}\].

**Fluorescence spectroscopy**

Titrations of PP2-A1-His$_6$ with N-N’-N”triacylchitotriose (Sigma) were performed by intrinsic fluorescence spectroscopy. We added 0.7 μg aliquots of N-N’-N”triacylchitotriose (C = 0.138 mg.mL$^{-1}$) stepwise to the protein solution (C = 0.175 mg.mL$^{-1}$) to obtain a protein to sugar molar ratio of 0 to 10. PP2-A1-His$_6$ was dissolved in “universal” buffer solution (Britton and Robinson type) at pH 8. “Universal” buffer contained citric acid, KH$_2$PO$_4$, H$_3$BO$_4$ and diethylbarbituric acid, and its pH was adjusted by titrating the solution with 0.2 M NaOH (Johnson and Lindsey, 1939; Britton, 1955). Intrinsic fluorescence spectra were recorded on a Fluoromax-Spex (Jobin-Yvon, Longjumeau, France) at 25°C, using an excitation wavelength of 290 nm. Spectra were collected with a step resolution of 1 nm and emission and excitation slits fixed at 5 nm. All spectra were corrected for the solvent (buffer in the presence of N-N’-N”-triacylchitotriose) in the 300-450 nm regions, using a 1-cm path length. Spectra were recorded in duplicate. The fluorescence intensity corresponding to sugar/protein complex formation during titration was used to calculate binding parameters $K_d$ (dissociation constant) and $n$ (number of sites) by the fitting of a Scatchard regression line (Scatchard, 1949).

In this case, the Scatchard equation was applied assuming a non cooperative model with independent and non identical sites. The equation used in this case is given by:

$$\frac{S_b}{S_f} = \frac{-S_b}{K_d} + \frac{nP_t}{K_d}$$

where $S_b$ is the bound sugar concentration (M), $S_f$ the free sugar concentration (M) and $P_t$ the total protein concentration (M). The $S_b/P_t$ ratio was given by the $F-F_0/F_{max}-F_0$ data values corresponding to the fluorescence of the sugar/protein complex. The free sugar concentration $S_f$ was deduced from $S_b/P_t$ values and total sugar concentration. The Scatchard representation displays $S_b/S_fP_t$ as a function of $S_b/P_t$.

**Sequence analysis**
Searches for motifs and sites of posttranslational modifications, such as glycosylation and phosphorylation, were carried out with Prosite (http://expasy.org/prosite; Sigrist et al., 2002). Multiple alignments were established with Multalin (http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html). We searched for lectin domains with Prosite and Pfam (http://pfam.sanger.ac.uk/). The accession numbers for the PP2s analyzed in this work were AtPP2-A1: O81865; CbmLec26: CAA78979; CbpLec26: CAA80364; CsLec26: AAM82557; CsLec17: AAM77344; CmmLec17: AAM77341; CmmLec26: AAM74921; AgPP2-1: AAM62132; Nictaba: AAK84134. The nucleotide sequence accession number for PP2-A1 is At4g19840.

**Phloem sap exudation**

Phloem sap was obtained by EDTA-facilitated exudation (King and Zeevaart, 1974). Four- to six-week-old plants were grown in a greenhouse (16 h photoperiod). Mature rosette leaves were cut at the base of the petiole and a second cut was made next to the first cut, in exudation buffer (potassium phosphate buffer 50 mM pH 7.6, EDTA 10 mM). The tip of the petiole was placed in 80 µL exudation buffer in a 2 mL Eppendorf tube. Standard exudation experiments were carried out for 6 h, in the dark, in a closed box containing wet paper to limit transpiration and to maintain humidity, in growth cabinets (20°C; 75% relative humidity). Samples were pooled and concentrated on centrifugal filter devices with a cutoff of 10 kD (Centricon, Millipore). The sugars in the phloem sap exudates were analyzed in an enzymatic assay based on a commercial kit for determining sucrose and hexoses (Sucrose/D-Glucose/D-Fructose kit RBIOPHARM).

**Analysis of phloem sap proteins by enzymatic glucosidase treatments**

For PNGaseF enzymatic treatment, the content of exudate samples was adjusted such that the exudate was essentially in PNGaseF reaction buffer (50 mM phosphate buffer pH 7.5, SDS 0.1%, 50 mM β-mercaptoethanol) and the samples were then boiled for 5 minutes. We added 0.7% Triton X-100 and 5 units of PNGase F (Sigma) and incubated the mixture overnight at 37°C. For β-N-acetylglucosaminidase (GDase) enzymatic treatment, the content of the exudate samples was adjusted to the GDase reaction buffer (50 mM citrate buffer pH 5, 100 mM NaCl, 0.01% BSA). We then added 5 mM N-acetyl β D glucosaminide and 0.01 units of GDase (Sigma). The mixture was incubated overnight at 25°C. Proteins were analyzed by SDS–PAGE in a 12% polyacrylamide gel. Proteins were visualized by staining with SYPRO Ruby (BioRad) for visualization at 580 nm with a phosphoimager, or blotted onto
nitrocellulose or polyvinylidene fluoride (PVDF) membranes (BioRad) for far-western immunodetection.

**Far-western blots**

For protein overlay experiments (far-western blots), the membranes onto which the proteins had been transferred were blocked by incubation for 1 h with 5% nonfat milk powder in PBS supplemented with 0.1% Tween 20 (PBS-T). They were then incubated for 1 h at room temperature with the recombinant PP2A1-His₆ (10 μg.mL⁻¹) in PBS-T. The recombinant protein was detected by incubation for 1 h at room temperature in PBS-T with a commercial peroxidase-conjugated mouse monoclonal antibody against His (Roche, 50 mU.mL⁻¹, 1/1000°), for the detection of histidine-tagged recombinant proteins. Peroxidase activity was detected with the Immun-Star™ HRP chemiluminescence kit (BioRad). The experiments were repeated five times.

**Analysis of phloem sap proteins by LC MS/MS**

Phloem sap proteins were separated by SDS-PAGE in 12% polyacrylamide gels. The gel was stained with Sypro Ruby (BioRad) and the two main bands were excised (2 mm in width) and analyzed by the PAPPSO platform (http://moulon.inra.fr/PAPPSO/analyses.html). The proteins were subjected to in-gel digestion with the Progest system (Genomic Solution) as follows. Gel pieces were washed in successive baths of (i) 10% acetic acid 40% ethanol and (ii) 100% acetonitrile (ACN). They were then washed in successive baths of (i) 25 mM NH₄CO₃ and (ii) 100% ACN. A volume of 40 μl was used for each bath. Gel pieces were then incubated in 40 μl of 10 mM DTT in 25 mM NH₄CO₃, for 30 minutes at 55°C and in 30 μL of 50 mM iodoacetamide in 25 mM NH₄CO₃, for 45 minutes at room temperature for cysteine reduction and alkylation, respectively. Proteins were subsequently digested for 6 h at 37°C with 20 μL of 125 ng of modified trypsin (Promega) dissolved in 20% methanol and 20 mM NH₄CO₃ per gel piece. The peptides were extracted successively with (i) 20 μL of 0.5% trifluoroacetic acid (TFA), 50% ACN and (ii) 20 μL of 100% ACN. The resulting peptide extracts were dried in a vacuum centrifuge and suspended in 25 μl of 0.08% TFA, 2% ACN for LC-MS/MS analysis. LC-MS/MS was performed on an Ultimate 3000 LC system (Dionex, Voisins le Bretonneux, France) connected to an LTQ Orbitrap mass spectrometer (Thermo Fisher, USA), with a nanoelectrospray ion source. Tryptic peptide mixtures were loaded onto the Pepmap C18 (0.3 x 5 mm, 100 Å, 5 μm; Dionex) at a flow rate of 20 μL.min⁻¹. After 4 minutes, the precolumn was connected to the Pepmap C18 (0.075 x 15 cm, 100 Å, 3
µm) separating nanocolumn and a linear gradient was established, from 2 to 36% of buffer B (0.1% formic acid, 80% ACN) in buffer A (0.1% formic acid, 2% ACN), at a flow rate of 300 nL.min⁻¹ over 50 minutes. Ionization was performed at the liquid junction, with a spray voltage of 1.3 kV applied to an uncoated capillary probe (PicoTip EMITER 10 µm tip ID; New Objective, USA). Peptide ions were automatically analyzed by the data-dependent method, as follows: full MS scan (m/z 300-1600) on an Orbitrap analyzer and MS/MS on the four most abundant precursors, with the LTQ linear ion trap. In this study, only +2 and +3 charged peptides were subjected to MS/MS, with an exclusion window of 1.5 minutes, using classical peptide fragmentation parameters: Qz = 0.22, activation time = 50 ms, collision energy = 35%.

The raw data generated by the LTQ-Orbitrap mass spectrometer were first converted into an mzXML file with ReADW (http://sashimi.sourceforge.net). Proteins were then identified with X!Tandem software 1 (X!Tandem tornado 2008.02.01.3, http://www.thegpm.org) used against an Arabidopsis protein database associated with a proteomic contaminant database. The X!Tandem search parameters were: trypsin specificity with a tolerance of one missed cleavage, fixed alkylation of cysteine and variable oxidation of methionine. The mass tolerance was fixed to 10 ppm for precursor ions and 0.5 D for fragment ions. The final search results were filtered with a multiple threshold filter applied at the protein level and consisting of the following criteria: protein identified with a minimum of nine different peptide sequences detected with a peptide E-value < 0.05.

3D modeling of PP2-A1
We carried out protein analysis with PROSAL, which provides links to a full range of web analyses for protein primary sequences (http://xray.bmc.uu.se/sbnet/prosal.html). Tertiary structure was predicted by searching for similarity to proteins with structures solved with the CBS @tome2 metaserver (http://abcis.cbs.cnrs.fr/AT2/meta.html). This analysis includes the TITO program (Labesse and Mornon, 1998), which can align sequences when the level of sequence identity between a query sequence and a 3D structure is low (10-30%). The results were validated by searches with other metaservers, such as BioInfo Bank (http://meta.bioinfo.pl/3djury.pl), using the 3D-Jury method (Ginalska et al., 2003; von Grothuss et al., 2003), Fold Recognition Server (version 2.6.0) (http://www.sbg.bio.ic.ac.uk/servers/3dpssm/) with the 3D-PSSM program (Kelley et al., 2000) and the PHYRE web server (http://www.sbg.bio.ic.ac.uk/phyre/html/index.html). A 3D
model was prepared with MacPyMOL software. The accession numbers for CBMs are as follows: for the Cbm4 of the laminarase 16a (lam16a) from the bacterium *Thermotoga maritima*: PDB id: 1gui1; Genbank Id: Q9WXN1; for the Cbm 22-2 of the endo-1,4-beta-xylanase Y (Xyn10B) from *Clostridium thermocellum*: PDB id: 1H6Y.

**Aphid artificial diet bioassay**

The stock population of *A. pisum* (clone LL01) was maintained in the laboratory on broad beans, as previously described (Febvay *et al.*, 1988). The stock population of *M. persicae* (“green” LC05 clone originating from INRA Colmar, and “pink” L06 clone collected on pepper in Lyon) was maintained on pepper. For both aphid species, the experiments were conducted on neonate aphid nymphs reared on aphid diet formulation AP3 (Febvay *et al.*, 1988). In addition to microelements and macroelements, this diet contains 584 mM sucrose and 260 mmol.L\(^{-1}\) amino acids. These conditions are typical of the sugar and nitrogen content of phloem sap. The medium was supplemented with various concentrations, from 0 to 1000 µg.mL\(^{-1}\) of recombinant PP2-A1-His\(_6\) (at a concentration of 0.01, 16, 32, 64, 125, 250, 500 or 1000 µg.mL\(^{-1}\) for biological replicate 1 and 0.01, 32, 125, or 500 µg.mL\(^{-1}\) for biological replicate 2 with *A. pisum*; 28, 83, 250 or 750 µg.mL\(^{-1}\) for experiments with *M. persicae* L06 and LC05 clones for biological replicate 1 and 0.01, 50, 100, 200, 300 and 500 for biological replicate 2), casein (Sigma) used as a negative control or wheat germ aggluninin (WGA, Sigma L9640) used as a control GlcNAc lectin (at a concentration of 28, 83, 250 or 750 µg.mL\(^{-1}\)). These concentrations are typical of those generally used in lectin bioassays in aphids (Rahbé and Febvay, 1993). Ampicillin (50 µg.mL\(^{-1}\)), an antibiotic harmless to the aphid primary symbiont *Buchnera*, was added to the medium to prevent microbial contamination. For each concentration of recombinant protein, we transferred 25 to 40 neonate nymphs directly onto artificial diet medium, on which they were maintained, at 20°C, for seven days. For the control proteins, casein and WGA, 18 neonate nymphs were used per concentration. The biological performance of PP2-A1-His\(_6\) was estimated by assessing mortality during development and by weighing aphids after the adult molt (measured on a 1 µg sensitive Setaram balance for *M. persicae*, and a 10 µg sensitive balance for *A. pisum*). All experiments were carried out at 20°C, 75% relative humidity, with a photoperiod of 16 h daylight/8 h darkness. The results were statistically analyzed, using ANOVA and common post-hoc tests, including Dunnet’s tests (JMP Software v8, SAS Institute Cary NC USA). In these experiments, aphid behavior was evaluated by scoring aphid settlement, measured after 24 h by counting the number of aphids successfully feeding on the artificial medium.
The choice test was carried out as described (Chen et al., 1996). Briefly, six neonate aphids were deposited in cylindrical boxes containing two diet tubes (one containing aphid diet medium and one containing aphid diet medium supplemented with PP2-A1-His$_6$), which were left overnight in a dark box for 16 h, before scoring for each set of conditions. The experiment was repeated at least 18 times for each choice situation. Two concentrations of PP2-A1-His$_6$ were tested: 200 and 500 µg.mL$^{-1}$. For each set of conditions, the phagostimulation index, I$_x$, was calculated as follows: I$_x$=T-C/T+C, with T=number of aphids on medium supplemented with PP2-A1-His$_6$ (test) and C=number of aphids on control medium. The results were analyzed with a non parametric paired-data method (Wilcoxon’s signed rank test, JMP software).

**Acknowledgments**

We thank H. Rogniaux and A. Geairon (UR1268, BIBS Platform, Nantes) for the mass spectrometry analyses, G. Duport and M. Pichon (INSA-INRA, Lyon) for the diet experiments on aphids, J.P. Compoint (UR1268 Nantes) for fruitful discussions on the PP2-A1 purification procedure and Dr M. Gonneau (LCB, Versailles) and Dr A. Guillot (UR477, PAPPSO Platform Jouy en Josas) for LC MS/MS analyses. We thank T. Lemaître for discussions and the Consortium for Functional Glycomics and Dr. D. Smith for providing the resources and carrying out the glycan microarray analyses. We thank S. Bouhallab (INRA, Rennes) for help with fluorescence data processing and Dr. G. Labesse for help with PP2-A1 3D structure prediction. We thank OUEST-GENOPOLE$^\text{®}$ for contributing to the recombinant protein work.
Literature Cited:


Allen AK (1979) A lectin from the exudate of the fruit of the vegetable marrow (Cucurbita pepo) that has a specificity for beta-1,4-linked N-acetylglucosamine oligosaccharides. Biochem J 183: 133-137


Balachandran S, Xiang Y, Schobert C, Thompson GA, Lucas WJ (1997) Phloem sap proteins from Cucurbita maxima and Ricinus communis have the capacity to traffic cell to cell through plasmodesmata. Proc Natl Acad Sci U S A 94: 14150-14155


Carlini CR, Grossi-de-Sa MF (2002) Plant toxic proteins with insecticidal properties. A review on their potentialities as bioinsecticides. Toxicon 40: 1515-1539


Vimr E, Lichtensteiger C (2002) To sialylate, or not to sialylate: that is the question. Trends Microbiol 10: 254-257


Figure Legends

Figure 1. Binding of recombinant PP2-A1-His$_6$ to a chitin affinity column
SDS-PAGE in a 15% polyacrylamide gel of the various fractions obtained during purification on a chitin affinity column. Lane 1: Soluble fraction of the bacterial culture after PP2-A1-His$_6$ production in *E. coli*. Lane 2: Unbound fraction after column loading. Lanes 3 - 4: Fractions eluted after washing with extraction buffer (50 mM Tris, 200 mM NaCl, pH 8). Lanes 5 - 7: Fractions eluted from the column with successive NaOH treatments (0.05, 0.1 and 0.5 M NaOH). This elution revealed the main bound protein to be PP2-A1-His$_6$. Lane 8: For comparison, loading of 1 µg of purified recombinant PP2-A1-His$_6$ obtained after loading on a nickel affinity column.

Figure 2. Spectrum of PP2-A1-His$_6$ lectin binding to the glycan microarray
A: Binding of PP2-A1-His$_6$ to glycans, expressed in relative fluorescence units (RFU) of PP2-A1-His$_6$ (at a concentration of 50 µg. mL$^{-1}$). The glycans giving the strongest signal (quantified in relative fluorescence units ± SEM) are shown, as well as several reference control sugars. The upper panel shows the minimal glycan structure recognized for each glycan category. *: glycans with a specific binding response (P value < 0.01). #: numbers correspond to the name of the glycans (see below).

B: Representation of the glycan structures displaying a specific response. Numbers correspond to the individual glycans, as indicated below. Mannose is represented by black circles, GlcNAc by gray squares, Gal by white circles and Neu5Ac by gray diamonds. PP2-A1-His$_6$ had the highest affinity for structures containing GlcNAc$_{\beta1-4}$GlcNAc$_{\beta1-4}$GlcNAc$_{\beta1-4}$ and Man$_{\alpha1-3}$(Man$_{\alpha1-6}$)Man$_{\beta1-4}$GlcNAc$_{\beta1-4}$GlcNAc$_{\beta}$ (encircled). PP2-A1-His$_6$ also displayed significant affinity for 9NacNeu5Ac$_{\alpha}$.

The names of the glycan structures represented in A and B are as follows: 1: Man$_{\alpha1-2}$Man$_{\alpha1-6}$(Man$_{\alpha1-3}$)Man$_{\alpha1-6}$(Man$_{\alpha2}$Man$_{\alpha2}$Man$_{\alpha1-3}$)Man$_{\beta1-4}$GlcNAc$_{\beta1-4}$GlcNAc$_{\beta}$; 2: Man$_{\alpha1-6}$(Man$_{\alpha1-3}$)Man$_{\alpha1-6}$(Man$_{\alpha2}$Man$_{\alpha1-3}$)Man$_{\beta1-4}$GlcNAc$_{\beta1-4}$GlcNAc$_{\beta}$; 3: Man$_{\alpha1-6}$(Man$_{\alpha1-2}$Man$_{\alpha1-3}$)Man$_{\alpha1-6}$(Man$_{\alpha2}$Man$_{\alpha1-3}$)Man$_{\beta1-4}$GlcNAc$_{\beta1-4}$GlcNAc$_{\beta}$; 4: Man$_{\alpha1-6}$(Man$_{\alpha1-3}$)Man$_{\alpha1-6}$(Man$_{\alpha1-3}$)Man$_{\beta1-4}$GlcNAc$_{\beta1-4}$GlcNAc$_{\beta}$; 5: Man$_{\alpha1-3}$(Man$_{\alpha1-6}$)Man$_{\beta1-4}$GlcNAc$_{\beta1-4}$GlcNAc$_{\beta}$; 6: Gal$_{\beta1-4}$GlcNAc$_{\beta1-2}$Man$_{\alpha1-3}$(Man$_{\alpha1-6}$)Man$_{\beta1-4}$GlcNAc$_{\beta1-4}$GlcNAc$_{\beta}$; 7: Neu5Ac$_{\alpha2-6}$Gal$_{\beta1-4}$GlcNAc$_{\beta1-2}$Man$_{\alpha1-3}$(Man$_{\alpha1-6}$)Man$_{\beta1-4}$GlcNAc$_{\beta1-4}$GlcNAc$_{\beta}$; 8: GlcNAc$_{\beta1-2}$Man$_{\alpha1-3}$(GlcNAc$_{\beta1-2}$Man$_{\alpha1-6}$)Man$_{\beta1-4}$GlcNAc$_{\beta1-4}$(Fuc$_{\alpha1-6}$)GlcNAc$_{\beta}$; 9: GlcNAc$_{\beta1-4}$GlcNAc$_{\beta1-3}$(GlcNAc$_{\beta1-2}$Man$_{\alpha1-6}$)Man$_{\beta1-4}$GlcNAc$_{\beta1-4}$GlcNAc$_{\beta}$.

Figure 3. Relative affinity of the three main classes of oligosaccharides recognized by PP2-A1-His6

Responses, in relative fluorescence units (RFU) of the three minimal glycan skeletons for which PP2-A1-His6 had the highest affinity at the various protein concentrations tested (0.1, 1, 10, 50 and 100 μg.mL⁻¹). Error bars: SEM; n=6. Manα1-3(Manα1-6)Manβ1-4GlcNAcβ is represented by a discontinuous gray line, GlcNAcβ1-4GlcNAcβ1-4 by a black line and 9NAcNeu5Ac by a solid gray line. These three glycans present a highly specific binding response (T-test < 0.01) at concentrations of 10, 50 and 100 μg.mL⁻¹. For comparison, GlcNAcβ1-4GlcNAcβ1-4, which is poorly recognized by PP2-A1-His6, is also represented on the diagram.

Figure 4. Surface plasmon resonance analysis of the properties of PP2-A1 binding to chitotriose

A: Sensorgrams of CM5 surface (expressed as response units: RU)-immobilized PP2-A1-His6 with N,N’,N”triacetylchitotriose injected at various concentrations (0, 5, 10, 20, 50, 100, 150, 200, 400, 600, 800 and 1000 μg.mL⁻¹). Sensorgrams were obtained in triplicate and were highly reproducible.

B: Plots of relative response (expressed in resonance units: RU) as a function of N,N’,N”triacetylchitotriose concentration (μg.mL⁻¹) on lin-lin and log-log representations.

Figure 5. Analysis of the fluorescence spectrum of PP2-A1 in the presence of chitotriose

A: Fluorescence spectra of PP2-A1-His6 alone (−) and PP2-A1-His6 with N,N’,N”triacetylchitotriose (−;; molar ratio 10.7) in “universal” buffer, pH 8. The wavelength of maximum fluorescence emission is indicated by the vertical dotted line. Spectra were recorded in duplicate and were reproducible.

B: Curve for the titration of PP2-A1-His6 with N,N’,N”triacetylchitotriose.
C: Scatchard plot of the data displayed in (B) with the fitting of a linear regression line (—). See text for explanations.

Figure 6. Hypothetical model of the tertiary structure of PP2-A1
This model presents a predicted carbohydrate-binding domain of PP2-A1 consisting of parallel and anti-parallel β-strands, based on structural similarities with known CBMs.
A: The positions of secondary structure elements for PP2-A1 are indicated on a ribbon representation from residues 110 to 249: β-strand 1 (136-139); β-strand 2 (143-146); β-strand 3 (152-163); β-strand 4 (167-179); β-strand 5 (187-194); β-strand 6 (203-211); β-strand 7 (219-223); β-strand 8 (231-239); β-strand 9 (242-245). Loops (solid lines) and strands (arrows) were predicted by alignment with CBM4 and CBM22. Regions for which the alignments provided no appropriate prediction are indicated by dotted lines.
B: Position of conserved domains: domain A (97-162), domain B (169-186); domain C (196-210); domain D (219-244).

Figure 7. Analysis of the binding of PP2-A1 to phloem sap proteins
A: Sugar content of phloem sap exudate (nmole per mg fresh weight). Error bars: SEM n=6
B: SDS-PAGE analysis (12% polyacrylamide gel) of phloem sap exudate (lane 2), including comparison with the total protein extracted from leaf veins (lane 1). The asterisk and numbers (1-5) in lane 2 indicate the various 26 kD and 52 kD fractions collected for LC MS/MS analysis (fractions 1 to 5, as reported in supplementary materials S3). For more details, see supplementary materials S3.
C: SDS-PAGE analysis (12% polyacrylamide gel) of phloem sap exudate before treatment (lanes 1 & 4) and after treatment with PNGase (lanes 2 & 5) or GDase (lanes 3 & 6). The migration of PNGase is indicated by an asterisk and that of GDase by an open circle. Proteins were detected by staining (lanes 1-3) or by protein overlay, using PP2-A1-His₆ and peroxidase-conjugated anti–His₆ antibody (lanes 4-6).

Figure 8. Effect of PP2-A1-His₆ on the growth of the aphids A. pisum and M. persicae
A: Growth of neonate nymphs of *A. pisum* after feeding for 7 days on an artificial diet containing various concentrations of PP2-A1-His$_6$, in two independent biological replicates (Repl.1 and Repl.2). Values are expressed as fractions of the mean weight for aphids fed on medium with and without recombinant protein. Error bars: SEM Aphid nymph number n>30. For all graphs, means differing from the control at a 5% risk level are indicated (*; Dunnet’s test).

B: Growth of neonate nymphs of *M. persicae* (clone L06) after feeding for 7 days on an artificial diet containing various concentrations of PP2-A1-His$_6$. Two independent biological replicates were realized (Repl.1 and Repl.2). Values are expressed as fractions of the mean weight for aphids fed on medium with and without recombinant protein. Error bars: SEM Aphid nymph number n>30. A two-way ANOVA on standardized weight data (% of control) showed no clone effect (p=0.986), no interaction (p=0.806) and a significant dose effect (p<0.0001); * labels a mean significantly different from control (Dunnet’s test).

C: Growth of neonate nymphs of *M. persicae* after feeding for 7 days on an artificial diet containing various concentrations of casein or wheat germ aggluninin (WGA) used as negative control (casein) or GlcNAc lectin control (WGA) with limited toxicity. Values are expressed as fractions of the mean weight for aphids fed on medium with and without protein. Error bars: SEM Aphid nymph number n>18. No effect of control proteins was detected (ANOVA p=0.54 for casein and p=0.44 for WGA).
A

B

1  MSKKHCSELL  PNKMFQNRQDS  KYLIPVQKEA  PPVTTLPMKA  STVKSPLHCE  AILRDADPPPI

61  SLSSVNLSEQ  LRSGVFLKPK  KQIKYIVWDER  NSNCFMLFRAK  NLSITWSDDV  NYWTVFTEKE

121  SPNENVEAVG  LKNVCWLWDLT  GKFDRNRLTP  GIVYVVVFKV  KLEQFWXGWV  TRYNLXMLLR

181  NGKEKPQREKK  VSLRELPRYK  WVDVRRGGEFV  PEKSAAGEIT  FSMYHAAAV  WVKGLSLKGV

241  AIRPKQ