Running head: Marchantia polymorpha expresses a typical fungal lectin

Research area: Biochemical Processes and Macromolecular Structures

Corresponding author:

Els J.M. Van Damme,
Department of Molecular Biotechnology,
Lab. Biochemistry and Glycobiology,
Ghent University,
Coupure links 653, B-9000 Gent, Belgium

E-mail: ElsJMVanDamme@UGent.be;
Fax +32 92646219;
Tel. +32 92646086
The liverwort *Marchantia polymorpha* expresses orthologs of the fungal *Agaricus bisporus* agglutinin family

Willy J. Peumans¹, Elke Fouquaert², Alain Jauneau³, Pierre Rouge⁴, Nausicaä Lannoo², Hiroki Hamada⁵, Richard Alvarez⁶, Bart Devreese⁷, Els J.M. Van Damme²

¹Nievelveldweg 9, B-9310 Aalst, Belgium

²Department of Molecular Biotechnology, Lab. Biochemistry and Glycobiology, Ghent University, Coupure Links 653, 9000 Gent, Belgium

³IFR40-CNRS, Signalisation et Biotechnologie Végétales, 24 Chemin de Borde Rouge, 31326 Castanet Tolosan, France

⁴Surfaces cellulaires et Signalisation chez les Végétaux, UMR CNRS-UPS 5546, Pôle de Biotechnologies Végétales, 24 Chemin de Borde Rouge, B.P. 17, 31326 Castanet-Tolosan, France

⁵Department of Applied Science, Okayama University of Science, 1-1 Ridai-Cho, Okayama 700-0005, Japan

⁶Department of Biochemistry and Molecular Biology, University of Oklahoma, Health Sciences Center, Oklahoma City, Oklahoma 73104

⁷Laboratory for Protein Biochemistry and Protein Engineering, Ghent University, KL Ledeganckstraat 35, 9000 Gent, Belgium
Footnotes

This research was supported by the Fund for Scientific Research-Flanders (Project No. G.0201.04) and the Research Council of Ghent University. The glycan array analysis was conducted by the Protein-Glycan Interaction Core H of The Consortium for Functional Glycomics funded by the National Institute of General Medical Sciences grant GM62116.

Corresponding author: ElsJM.VanDamme@UGent.be; Fax : +32 92646219

Key words: Agaricus bisporus agglutinin, Marchantia polymorpha, liverwort, Tortula ruralis, subcellular location
Abstract
A lectin different from the previously described mannose-binding agglutinins has been isolated from the liverwort *Marchantia polymorpha*. Biochemical characterization of the purified lectin combined with the data from earlier transcriptome analyses demonstrated that the novel *M. polymorpha* agglutinin is not related to any of the known plant lectin families but closely resembles the *Agaricus bisporus* type lectins, which hitherto have been found exclusively in fungi. Immunolocalization studies confirmed that the lectin is exclusively associated with the plant cells ruling out the possibility of a fungal origin. Extensive screening of the publicly accessible databases confirmed that apart from fungi the occurrence of *A. bisporus* type lectins is confined to *M. polymorpha* and the moss *Tortula ruralis*. The expression of a typical fungal protein in a liverwort and a moss raises the question of the origin of the corresponding genes. Irrespective of the evolutionary origin, the presence of a functional *A. bisporus* lectin ortholog in *M. polymorpha* provides evidence for the expression of an additional carbohydrate binding domain in Viridiplantae.

Abbreviations

ABA: *Agaricus bisporus* agglutinin; EA: elisa array; GalNAc: N-acetylgalactosamine; GlcNAc: N-acetylglucosamine; GNA: *Galanthus nivalis* agglutinin; MarpoABA: *Marchantia polymorpha* ABA ortholog; ORF: open reading frame; PA: printed array; T-antigen: Thomsen-Friedenreich antigen (Galβ1,3GalNAcα1-O-Ser); TorruABA: *Tortula ruralis* ABA ortholog; UTR: untranslated region
INTRODUCTION

Modern plant biology research provided a fairly detailed overview of the occurrence and identity of carbohydrate-binding proteins in plants. Apart from a few exceptions, all of the approximately 500 currently known plant lectins can be classified into seven families of structurally and evolutionary related proteins (for reviews see Van Damme et al., 1998; Van Damme et al., 2004). Three lectin families, namely the amaranthins, the jacalin-related lectins and the Cucurbitaceae phloem lectins are apparently confined to plants. Until recently, legume lectins also were considered typical plant proteins because no similar lectins were found in other organisms. However, structural analyses (Velloso et al., 2003) clearly demonstrated that the so-called ERGIC (ER-Golgi intermediate compartment) proteins, which are absent from plants but are found in animals and fungi (Dodd and Drickamer, 2001) share a common ancestor and accordingly should be considered homologs of the legume lectins. In contrast, homologs of the monocot mannose-binding lectins have been isolated from a fish (Fugu rubripes) (Tsutsui et al., 2003) and from the slimemold Dictyostelium discoideum (Jung et al., 1996). Hevein domains also have been identified in several proteins from metazoa (e.g. Caenorhabditis elegans) and fungi, whereas domains corresponding to the sugar-binding chain of type-2 ribosome-inactivating proteins are widespread among both prokaryotes and eukaryotes. These observations raise some important questions about the evolutionary origin of plant lectins. Apart from the ricin-B domain and to a lesser extent the hevein domain, there is little evidence for a classical vertical inheritance from a common prokaryotic ancestor. In addition, the evolution of the lectin families within the Viridiplantae is still poorly understood because apart from a few exceptions, all plant lectins were identified in flowering plants. For the time being, the documented occurrence of lectins (with a known sequence) outside the flowering plants is limited to jacalin-related lectins in a cycad (Cycas revoluta) (Yagi et al., 2002) and a fern Phlebodium aureum (Tateno et al., 2003), and a monocot mannose-binding lectin in the gymnosperm Taxus media (Kai et al., 2004) and the liverwort Marchantia polymorpha (Peumans et al., 2002).
Here we present evidence that the liverwort *Marchantia polymorpha* and the moss *Tortula ruralis* express proteins that can be considered orthologs of the *Agaricus bisporus* lectin family. Biochemical characterization of the purified lectin confirmed that some of the orthologs expressed by *M. polymorpha* are fully active agglutinins with a carbohydrate-binding specificity similar to that of the *A. bisporus* agglutinin (ABA). Immunolocalization studies clearly demonstrated that the lectin is exclusively associated with the *M. polymorpha* cells ruling out the possibility that the protein is produced by a contaminating fungus. These findings not only provide firm evidence for the expression in plants of a lectin type that hitherto has been found exclusively in fungi but also add a novel family to the list of previously identified plant lectin families. No other plant sequences could be identified that encode similar proteins indicating that within the Viridiplantae the occurrence of the *A. bisporus* type lectin is confined to lower plants. The impact of these novel findings on the unraveling of the evolution of lectins in general and plant lectins in particular is discussed.
RESULTS AND DISCUSSION

ABA orthologs are found in the liverwort *Marchantia polymorpha* and the moss *Tortula ruralis*

A screening of the publicly accessible databases for the occurrence in lower plants of homologs/orthologs of lectins found in flowering plants and fungi led to the identification of expressed sequence tags (ESTs) from the liverwort *M. polymorpha* and the moss *T. ruralis* that encode proteins with a marked sequence similarity (>40% and >70% sequence identity and similarity, respectively) to the *A. bisporus* agglutinin. To corroborate the possible occurrence of other non-fungal ABA-related lectins, protein and DNA databases were extensively searched using the sequences of all known fungal and *M. polymorpha* and *T. ruralis* ABA orthologs as a query. No additional non-fungal ABA-related proteins could be retrieved.

**Immature female sex organs of the liverwort *Marchantia polymorpha* express multiple ABA orthologs**

A total number of 17 ESTs encoding proteins similar to ABA were identified in the *M. polymorpha* EST database (containing a total number of 1,415 EST entries). Due to high overall sequence identity/similarity the *M. polymorpha* proteins are considered ‘ABA orthologs’. All 17 ESTs were retrieved from a library prepared from immature female sex organs (Nagai et al., 1999). None was found in the library prepared from immature male sex organs (Nishiyama et al., 2000). In the library from immature female sex organs the ESTs encoding ABA-orthologs represent 1.75% of the library total (970). The occurrence of ESTs encoding proteins similar to ABA was already noticed by Nagai et al. (1999) in their original report of the analysis of the ESTs from *M. polymorpha* immature female sexual organs. According to this report these ESTs were grouped in 4 contigs (# 3, 17, 51 and 55). However, a more detailed analysis of the nucleotide sequences revealed that these
17 ESTs should be assembled into five different contigs (see Supplemental Figs. S1-S3), which according to the length of the encoded polypeptide can be subdivided into two subgroups. The first subgroup comprises three contigs (referred to as MarpoABA1a, MarpoABA1b and MarpoABA1c, respectively, see Supplemental Figs. S1-S2) encoding an ABA ortholog of 140 amino acid residues (Fig. 1). The second subgroup comprises two contigs (referred to as MarpoABA2a and MarpoABA2b, respectively) encoding an ABA ortholog of 142 amino acid residues (see Supplemental Fig. S3).

Analyses of the complete deduced amino acid sequences allowed calculating the major physicochemical parameters of the corresponding proteins (see Supplemental Table S1) and, in addition, indicated that all *M. polymorpha* ABA orthologs are synthesized without a signal peptide.

**Rehydrating gametophytes of the desiccation-tolerant bryophyte *Tortula ruralis* express a single ABA ortholog**

Two ESTs encoding a putative ortholog of ABA were identified in the *T. ruralis* EST database (containing a total number of 9,306 EST entries). Both ESTs have an identical nucleotide sequence and encode a protein of 144 residues (further referred to as TorruABA). The deduced sequence of TorruABA shares approximately 50% and 80% sequence identity and similarity, respectively, with the *M. polymorpha* ABA orthologs, and lacks a signal peptide. TorruABA aligns best with the MarpoABA1 group (see Supplemental Fig. S4). The major physicochemical parameters of TorruABA could be calculated from the complete deduced amino sequence (see Supplemental Table S1).

**PCR amplification and sequence analysis of genomic sequences encoding *M. polymorpha* ABA orthologs**
To confirm the presence of genomic sequences corresponding to the contigs assembled from the deposited ESTs DNA was extracted from young female sexual organs of *M. polymorpha* and genomic sequences encoding the ABA orthologs amplified by PCR. Analysis of the PCR fragments confirmed the presence in the genome of sequences corresponding to the ORFs of MarpoABA1a, MarpoABA1b/MarpoABA1c, and MarpoABA2a and MarpoABA2b. In addition, a PCR fragment with a novel sequence representing a third subgroup (referred to as MarpoABA3a) (see Supplemental Fig. S5) encoding a lectin polypeptide of 146 amino acid residues was obtained (Fig. 1). Though it can not be precluded that some genomic sequences were not amplified or detected, the results of the PCR amplification experiments confirm that -as could already be inferred from the EST data- the expression of ABA orthologs in *M. polymorpha* is controlled by a gene family.

PCR amplification experiments were also set up with DNA isolated from an axenic culture of *M. polymorpha*. Analysis of the amplified fragments confirmed that the DNA of the axenically grown *M. polymorpha* cells contains the same lectin sequences as the DNA from the field grown sample, which strongly indicated that the lectin is synthesized by the plant cells and is not derived from a possible fungal contaminant.

**A previously characterized *Marchantia polymorpha* lectin corresponds to MarpoABA2**

The first lectin to be characterized from a lower plant was isolated from the liverwort *M. polymorpha*. Using a combination of classical protein purification techniques Adam and Becker (1993) purified a lectin from *M. polymorpha* gametophytes and characterized the protein in some detail. The lectin behaved as a monomeric protein with a Mr of 16,135 (as determined by mass spectrometry) and is not glycosylated. Agglutination assays further indicated that the lectin agglutinates erythrocytes of different mammalia, and that this agglutination activity can not be inhibited by any simple sugar but only by complex carbohydrate structures. Both the Mr of the polypeptide and the specificity of the lectin are
reminiscent to the ABA type fungal lectin. Moreover, the molecular mass as determined by mass spectrometry is almost identical to the values calculated from the deduced sequences of the *M. polymorpha* ABA (Table S1). It is very likely, therefore, that the *M. polymorpha* lectin described by Adam and Becker (1993) corresponds to an ABA ortholog.

To corroborate the identity, the lectin was reisolated and characterized. The purified protein yielded a single polypeptide band of approximately 15 kDa upon SDS-PAGE (Results not shown) and eluted with an apparent Mr of approximately 30 kDa upon gel filtration on a Superose 12 column, indicating that the native lectin is a dimeric protein.

Mass spectrometry of the purified lectin yielded a main peak of 16102.0 +/- 0.55 Da and three minor peaks with molecular weights of 16118.5 (+/-1.22) Da, 16133.0 (+/- 1.65) Da and 16350.0 (+/-0.63) Da, respectively (see Supplemental Fig. S6). Since the protein was N-terminally blocked (as was already reported by Adam and Becker, 1993), partial sequencing was achieved by mass spectrometry. The purified lectin sample was digested with trypsin and the fragments analyzed by MS and MS/MS. Comparison of the peptide mass fingerprint with the calculated masses of the tryptic peptides of the different *M. polymorpha* ABA orthologs indicated that the major component of the purified lectin corresponds to MarpoABA2b. MarpoABA2a, which is identical to MarpoABA2b except that residue 136 is a Tyr instead of a Phe (and hence has a mass that is 16 Da higher), most probably corresponds to the second peak (16118.5 Da) in the ESI spectrum (which differs by 16 Da from the first peak). The peak at 16133.0 Da in the ESI spectrum might represent an oxidized form. No match could be found for the 16350.0 Da component. Possibly, this peak is due to the presence of a yet unidentified molecular species.

Analysis of the tryptic peptides by MS/MS yielded no indications for the presence of the (expressed) MarpoABA1 group. This obvious absence from the affinity-purified lectin preparation is in good agreement with the results of the modeling/docking experiments (see below), which indicated that the major carbohydrate binding site of MarpoABA1 isoforms is (in contrast to that of the MarpoABA2 isoforms) not functional so that they are not capable of binding the immobilized ligands.
The data presented in this section demonstrate that the lectin isolated from *M. polymorpha* corresponds to the MarpoABA2 orthologs. Though the *M. polymorpha* ABA ortholog can not be considered a novel type of lectin, its identification provides firm evidence for the expression of a typical fungal lectin in a (lower) plant.

**Carbohydrate-binding specificity of the *M. polymorpha* ABA ortholog**

Hapten inhibition assays indicated that the agglutination activity of the *M. polymorpha* ABA ortholog is not sensitive to any common simple mono- or oligo-saccharide but is completely inhibited by low concentrations (in the µg/mL range) of animal glycoproteins like asialomucin and asialofetuin. These observations confirm the results of the preliminary specificity studies reported by Adam and Becker (1993) but do not allow comparing the specificity of the *M. polymorpha* lectin to that of its fungal counterparts. Therefore the specificity of the *M. polymorpha* ABA orthologs was corroborated in more detail using the high performance glycan array systems developed by the Consortium for Functional Glycomics (see Supplemental Fig. S7). In a screening with the printed array platform (PA) the lectin strongly interacted with 9 glycans (Table I). All 9 glycans contain β-D-Gal or α-GalNAc. Unsubstituted β-D-Gal and α-GalNAc were nearly equally reactive. Addition of β1-3 linked Gal to α-GalNAc (giving rise to Galβ1-3GalNAcα- or Thomson Friedenreich or T antigen) slightly increased the interaction with the lectin. Substitution of T at O3 with sulfate, β1-3 linked GlcNAc or α2-3 linked Neu5Ac had no negative effect indicating that the lectin tolerates substitution. The same holds true for a substitution of the α-GalNAc at O6 with sulfate or α2-6 linked Neu5Ac. In contrast, substitution of α-GalNAc at O6 with β1-6 linked NeuAc reduced the binding with approximately half. The results of the elisa array (EA) were less clearcut. Of the 6 glycans common to both the printed and the elisa array only two, namely T-antigen and Sulfo-T showed binding on the elisa array. These obvious differences in reactivity most likely rely on differences in both the construction of the two glycan arrays and the binding assay itself. First, the PA has a higher density (though unknown amount) of glycans in each spot than the EA. Second, the lectin
concentration used in the screening was 10 times higher in the PA assay than in the EA assay (300 µg mL\(^{-1}\) and 30 µg mL\(^{-1}\), respectively). As a result the effect of a possible multivalency of the lectin-glycan interactions is much higher in the PA assay than in the EA assay. The fact that a number of glycan epitopes that do not bind in the EA assay exhibit a clear reaction in the PA assay strongly indicates that the presence of high-density polyvalent glycotopes strongly enhances the affinity of the lectin. Both the identity of the reactive glycans and the apparent preferential binding to multivalent epitopes indicate that the *M. polymorpha* lectin closely resembles ABA for what concerns its specificity (Wu et al., 2003).

**Immunolocalization confirms that the *M. polymorpha* ABA orthologs are synthesized by the plant cells and do not originate from a contaminating fungus**

The presence in a liverwort of ABA orthologs raised the question whether the lectin and the corresponding ESTs/genes are possibly derived from a contaminating fungus. To address this critical question the lectin was localized in axenically grown callus and field grown thallus tissue by confocal microscopy using highly specific antibodies. Fig. 2 clearly demonstrates that in both callus and thallus tissue sections immunolabeling is exclusively associated with the plant cells. A closer examination of the micrographs of the thallus sections indicates a stronger labeling in the upper and lower epidermal cell layers, the rhizoids and the small parenchyma cells located below the upper epidermal cell layer as compared to the large parenchyma cells. No signal was detected in the control samples (not shown). These findings unambiguously demonstrate that the ABA orthologs are synthesized by the plant cells and do not originate from a fungal contaminant. Moreover, there is no evidence whatsoever for the presence of fungal hyphae in the sections from field grown thalli.

The fluorescence patterns shown in Fig. 2 give no decisive answer that the *M. polymorpha* ABA ortholog is like the fungal ABA ortholog from *A. oligospora* (Rosén et al., 1997) confined to the cytoplasmic/nuclear compartment. On a first sight the lectin seems to be
associated with the cell wall of the *M. polymorpha* cells. However, further studies are required to corroborate the exact location of the lectin in the plant cells.

**Molecular modeling of the structure of the *M. polymorpha* and *T. ruralis* ABA orthologs**

To check whether the binding sites of the *M. polymorpha* and *T. ruralis* ABA orthologs can be functional their overall fold and three-dimensional structure was determined by molecular modeling using the atomic coordinates of ABA as a model. The modeled MarpoABA1a and TorruABA exhibit an overall three-dimensional fold similar to that found in ABA (Carrizo et al., 2005) (Fig. 3A-C).

In ABA the amino acid residues Ala29, Ser48 and Asn73, which form the main carbohydrate-binding site create a network of 5 hydrogen bonds with both the T-antigen (Galβ1,3GalNAc1α-O-Ser) and the T-antigen disaccharide (Galβ1,3GalNAc) (Fig. 3D) (Carrizo et al., 2005). An additional hydrogen bond mediated by a water molecule is formed between His72 and the GalNAc moiety of both the T-antigen and the T-antigen disaccharide. Another water-mediated hydrogen bond occurs between Arg107 and the GalNAc moiety of the T-antigen disaccharide. All these amino acid residues are strictly conserved in MarpoABA2a and MarpoABA2b. Accordingly, one can reasonably expect that MarpoABA2a and MarpoABA2b interact in a similar way with both the T-antigen and the T-antigen disaccharide as ABA itself. Docking experiments confirm that these interactions can take place (Fig. 3E). In addition, the results of the glycan array binding assays provide direct experimental evidence that a mixture of MarpoABA2a and MarpoABA2b interacts with the T-antigen disaccharide (Table I).

In contrast to MarpoABA2a/2b, MarpoABA1a, MarpoABA1b and MarpoABA1c as well as TorruABA lack the Ala29 residue and hence interact more weakly with the T-antigen. Due to the replacement of Asn73 by a Val MarpoABA3a is most probably unable to bind
the T-antigen (Table S2). These predictions are in good agreement with the absence of MarpoABA1a,b,c and MarpoABA3 isoforms from the affinity purified lectin preparation.

ABA contains besides its major T-antigen binding site a second (minor) site that interacts with GlcNAc (Carrizo et al., 2005; Nakamura-Tsuruta et al., 2006). This GlcNAc-binding site comprises the five amino acid residues Asp79, Ile80, Thr82, Arg103 and Tyr114 which establish a network of 8 hydrogen bonds with O3, O4, O5, O6 and O7 of GlcNAc, respectively (Fig. 3F). MarpoABA1a/1b/1c possess identical or similar residues at homologous positions and according to docking experiments interact with GlcNAc in a similar way as ABA (Fig. 3G). The replacement of Thr82 by a Val residue in MarpoABA1a/1b/1c has probably little effect on the binding of GlcNAc since both Thr and Val residues interact with O7 of GlcNAc through their N atom that is similarly positioned at the end of the conserved β7 sheet in all the lectins of the A. bisporus agglutinin family (Fig. 1). The replacement of the key residue Tyr114 (of ABA) by an Ile113 (MarpoABA2a/2b), a Met113 (MarpoABA3a) or a Leu113 (TorruABA) residue eliminates two stabilizing hydrogen bonds with O5 and O6 of GlcNAc, respectively, and hence severely hampers or even prevents the interaction of these lectins with GlcNAc. In summary, the results of the modeling/docking experiments (Table S2) indicate that MarpoABA2a/b possess a fully functional T-antigen and a weakly active GlcNAc-binding site whereas MarpoABA1a/b/c contain a fully active GlcNAc-binding sites but a T-antigen-binding with a reduced activity and MarpoABA3a possesses neither a functional T-antigen nor a GlcNAc-binding site. In TorruABA both binding sites most probably have a reduced activity. Though not conclusive, the results of the modeling/docking experiments have some predictive value with respect to the agglutinating activity of the different lectins. Thereby it should be taken into consideration that the agglutinating activity of ABA is mainly (or even exclusively) determined by the T-antigen-binding site (because no agglutination occurs in the presence of free T-antigen). Only MarpoABA2a and MarpoABA2b possess a lectin activity comparable to that of ABA. MarpoABA1a, MarpoABA1b, MarpoABA1c and TorruABA possibly display a (strongly) reduced lectin activity whereas MarpoABA3a is devoid of lectin activity.
Why does *M. polymorpha* express lectins with specificity towards the T-antigen disaccharide and GlcNAc?

The identification of ABA orthologs in *M. polymorpha* raises the question why this liverwort expresses cytoplasmic/nuclear lectins with specificity towards the T-antigen disaccharide. At present, the biological role of the fungal ABA orthologs is not fully understood. Studies with *Arthrobotrys oligospora* indicated that in this fungus the ABA-ortholog behaves as a cytoplasmic storage protein (Rosén et al., 1997), whereas a role as an insecticidal protein was attributed to the ABA ortholog from *Xeroomus chrysenteron* (Trigueros et al., 2003). Taking into consideration that all studied fungal ABA orthologs have a similar specificity the defensive role proposed for the *X. chrysenteron* agglutinin might also apply to other members of this lectin family. In this respect, the ABA-type fungal lectins can be regarded as functional homologs of the potent insecticidal T- and/or Tn-antigen binding plant lectins (like jacalin and the *Glechoma hederacea* lectin) (Czapla and Lang, 1990; Singh et al., 2006; Wang et al., 2003). However, neither of these two types of typical plant lectins has been found in *M. polymorpha*. Though purely speculative, one can imagine that the liverwort used a typical fungal protein as an alternative for the insecticidal lectins expressed in flowering plants. This reasoning does not apply to the defense against sucking insects because as has been demonstrated previously, *M. polymorpha* expresses fully active orthologs of the *Galanthus nivalis* agglutinin (GNA) (Peumans et al., 2002), which is a potent defense protein against sucking insects (like aphids and white flies). Possibly the simultaneous expression of both an ABA- and a GNA-type lectin offers *M. polymorpha* protection against a broad range of chewing and sucking insects.

**CONCLUSIONS**

*M. polymorpha* expresses functional orthologs of a lectin that hitherto was exclusively found in fungi. Construction of a phylogenetic tree from the available sequences not only confirms that the *M. polymorpha/T. ruralis* ABA orthologs belong to the same protein
family as the fungal ABA orthologs but also indicates that some fungal proteins are more closely related to the *M. polymorpha/T. ruralis* proteins than to other fungal ABA orthologs (Fig. 4, see Supplemental Fig. S8). The identification of homologous/orthologous lectins in both plants and fungi is not novel because representatives of e.g. the ricin-B family have already been isolated from flowering plants and fungi (as well as from numerous animals and prokaryotes). However, the finding that ABA related proteins are apparently confined to fungi and some lower plants is unexpected but highly relevant for what concerns the molecular evolution of lectins in general and plant lectins in particular. The apparent absence of homologous/orthologous genes from the genomes of all (sequenced) bacterial and animal genomes is difficult to explain, indeed, in terms of a classical vertical inheritance from a common (prokaryotic) ancestor. In addition, the question arises why no orthologs of the *M. polymorpha* and *T. ruralis* lectin genes are present in the genomes of embryophyta. Irrespective of the answer to this question, the identification of ABA orthologs in *M. polymorpha* and *T. ruralis* confirms the previously made observations (based on transcriptome analysis) that some lower plants express proteins that are absent from higher plants. Such observations were not only made for *M. polymorpha* and *T. ruralis* but also for the moss *Physcomitrella patens*. In the latter case only about 50% of the expressed protein genes could be matched to an Arabidopsis homolog (Rensing et al., 2002).
EXPERIMENTAL PROCEDURES

 Retrieval of sequences

Sequences encoding ABA-related proteins were retrieved by BLAST searches using the amino acid sequence of ABA as a query. In a later stage the deduced sequences of the ABA orthologs found in *M. polymorpha* and *T. ruralis* were used as a query. All retrieved EST sequences were analyzed individually. In the absence of complete ESTs, contigs were reconstructed from ESTs showing overlaps of at least 200 identical nucleotides. Searches were completed on December 15, 2005. The following databases were screened for the presence of EST and/or genomic sequences encoding plant orthologs of ABA: National Center for Biotechnology Information (NCBI: http://www.ncbi.nlm.nih.gov/), cosmoss.org (http://www.cosmoss.org/), Phytome (http://www.phytome.org/search.php), The Institute for Genomic Research (TIGR: http://tigrblast.tigr.org/tgi/) and Solanaceae Genomics network (http://www.sgn.cornell.edu/).

 Plant material

*M. polymorpha* L. thalli were collected locally in the beginning of December. Immediately after collection, the thalli were exhaustively rinsed with tap water to remove soil particles and organic debris trapped between the rhizoids. The washed thalli were further processed manually to remove the dead tissue at the proximal end. An axenic culture of *M. polymorpha* was obtained from Okayama University and maintained as described before (Ishihara et al., 2003).

 Isolation of the *M. polymorpha* ABA ortholog

Freshly harvested and processed thalli (500 g) were transferred into 5 L of a solution of 20 mM unbuffered 1,3 diaminopropane containing 0.1% thiourea and homogenized with a Waring blender. The homogenate was filtered through cheesecloth, centrifuged (3000 g for 15 min) and the supernatant filtered through filter paper (Whatman 3MM). The cleared filtrate was diluted with an equal volume of distilled water and loaded on a column of Q Fast Flow (Amersham Biosciences, Uppsala, Sweden; 5 cm x 5 cm; approximately 100 mL
bed volume) equilibrated with 20 mM unbuffered 1,3 diaminopropane. After passing the extract the column was washed with 500 mL of 20 mM unbuffered 1,3 diaminopropane and the bound proteins eluted with 250 mL of 0.5 M NaCl in 0.1 M Tris-HCl (pH 7.8). This partially purified protein fraction was adjusted to pH 7.5 with 1N HCl and loaded on a column of asialomucin-Sepharose 4B (2.6 cm x 5 cm; 25 mL bed volume) equilibrated with PBS. After loading the protein fraction, the column was washed with PBS until the A280 fell below 0.01, and the lectin eluted with 100 mL 20 mM unbuffered 1,3 diaminopropane. Since this lectin fraction was still strongly colored (because of the presence of some brown materials that were aspecifically bound to the affinity column) it was rechromatographed on the same affinity matrix to improve the purity. The pH of the first eluate was adjusted to 7.5 with 1 N acetic acid and solid NaCl added to a final concentration of 0.2 M. The brown precipitate that formed upon lowering the pH was removed by centrifugation (3,000 g for 15 min in 50 mL Falcon tubes) and the supernatant loaded on a small column (1.6 cm x 5 cm; approximately 10 mL bed volume) of asialomucin-Sepharose 4B. After washing with PBS until the A280 fell below 0.01 the bound lectin was desorbed with 20 mL 20 mM unbuffered 1,3 diaminopropane. To concentrate the affinity-purified lectin, the eluate was loaded on a small column (1 cm x 2 cm; 1.5 mL bed volume) of Q Fast Flow equilibrated with 20 mM unbuffered 1,3 diaminopropane and the lectin eluted with 3 mL of 0.4 M NaCl in 0.1 M Tris-HCl pH 7.8. This concentration step also improved the purity of the lectin because most of the impurities present in the lectin fraction after the second affinity chromatography did not elute with the lectin. Since the concentrated lectin solution was still slightly colored, a gel filtration chromatography step was added to the purification scheme. The fraction eluted from the Q Fast Flow column was chromatographed on a column (2.6 cm x 70 cm; approximately 350 mL bed volume) of Sephacryl 100 using 20 mM unbuffered 1,3 diaminopropane as a running buffer. Under these conditions the lectin eluted in a symmetrical peak well before the colored compounds. The fractions containing the lectin were pooled, dialyzed against water and lyophilized. Approximately 2 mg of pure lectin was obtained starting from 500 g of thalli.

Analytical techniques
The purified lectin was analyzed by SDS/PAGE in 15% (w/v) acrylamide gels as described by Laemmli (1970). Analytical gel filtration was performed on a Superose 12 column (Amersham Biosciences, Uppsala, Sweden) using phosphate buffered saline (PBS; 1.5 mM KH₂PO₄, 10 mM Na₂HPO₄, 3 mM KCl, 140 mM NaCl, pH 7.4) containing 0.2 M galactose as running buffer. Banana thaumatin-like protein (20 kDa) and β-glucanase (30 kDa) were used as molecular weight markers. Total neutral sugar was determined by the phenol/H₂SO₄ method with D-glucose as standard (Dubois et al., 1956).

For N-terminal amino acid sequencing purified proteins were separated by SDS-PAGE and electroblotted on a PVDF membrane. Polypeptides were excised from the blots and sequenced on a model 477A protein sequencer interfaced with a model 120A on-line analyzer (Applied Biosystems, Foster City CA, USA).

Mass spectrometric analysis of intact lectin
The molecular weight of the intact lectin was determined using electrospray ionization mass spectrometry on a Micromass Q-TOF I mass spectrometer (Waters, Milford, MA) equipped with an automated nano-electrospray source (Advion, Ithaca, NY). A solution of 5 pmol µL⁻¹ was prepared in 50 % acetonitrile/0.1 % formic acid in water of which 2 µL was loaded on the ESI chip and sprayed using a capillary voltage of 1100 V. Spectra were collected for 3 min with 1 s scans covering a m/z range from 500 to 2000 Da/e. The spectrum was processed using the Masslynx software delivered with the instrument.

Mass spectrometric identification of the lectin
Lectin polypeptides were excised from the gel and treated for tryptic digestion according to a protocol described previously (Devreese et al., 2002). After digestion, the supernatant was recovered, and the gel slices re-extracted twice with 50 µL of 60% acetonitrile/0.1% formic acid. The combined fractions were dried in a Speedvac concentrator (Thermo Savant, Holbrook, NY, USA).

The peptide mixture was applied on a 4700 Proteomics Analyzer, a MALDI TOF-TOF mass spectrometer (Applied Biosystems, Framingham, CA, USA), equipped with a Nd:YAG laser at 200 Hz rate. One 1 µL of the digest mixture (after reconstitution in 12 µL
0.1% formic acid) was applied, mixed with 1 µL 50 mM α-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 0.1% TFA on a MALDI target plate. Prior to analysis, the mass spectrometer was externally calibrated with a mixture of angiotensin I, Glu-fibrino-peptide B, adenocorticotrophic hormone (ACTH) (1–17), ACTH (18–39). For MS/MS experiments, the instrument was externally calibrated with fragments of Glu-fibrino-peptide. The mass spectrometric results were compared to the theoretical masses of the peptides encoded by the lectin genes.

**PCR amplification of genomic DNA fragments**

DNA was extracted from thalli, female sexual organs of *M. polymorpha* and axenically cultures callus cells using the protocol described by Stewart and Via (1993). Alternatively, DNA was extracted using the FastDNA Spin kit in an automatic homogenizer (FastPrep Instrument, MP Biomedicals and Qbiogene, Irvine, CA, USA) following the manufacturer’s recommendations. Genomic sequences encoding ABA orthologs were amplified by PCR using primers (i) derived from the 5’ and 3’ untranslated sequence or (ii) derived from the 5’ and 3’ ORF of ABA orthologs of *M. polymorpha*. The reaction mixture for amplification of genomic DNA sequences contained 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 100 mg L⁻¹ gelatin, 0.4 mM of each dNTP, 2.5 units of Taq polymerase (Invitrogen, Carlsbad, CA), 500 ng of genomic DNA and 20 µL of the appropriate primer mixtures (20 µM), in a 100 µL reaction volume. After denaturation of the DNA for 5 min at 95°C amplification was performed for 30 cycles through a regime of 15 sec template denaturation at 92°C followed by 30 sec primer annealing at 50°C and 1 min primer extension at 72°C. The PCR fragments were purified using Qiaquick PCR Purification kit (Qiagen, Hilden, Germany) and cloned in pCR2.1-TOPO cloning vector using the TOPO TA cloning kit from Invitrogen (Carlsbad, CA). Plasmids were isolated from purified single colonies on a miniprep scale using the alkaline lysis method (Mierendorf and Pfeffer, 1987) and sequenced by the dideoxy method (Sanger et al., 1977).

**Preparation of monospecific antibodies**
Polyclonal antibodies were raised against the *M. polymorpha* ABA orthologs in a female New Zealand White rabbit. The animal was injected subcutaneously with 1 mg of the purified lectin dissolved in PBS and emulsified in 1 mL of Freund’s complete adjuvant. Four booster injections with 0.2 mg lectin in 1 mL of PBS were given with 10-day intervals. Ten days after the final injection, blood was collected from an ear marginal vein and the crude serum prepared by standard techniques. Western blot analysis of crude extracts from *M. polymorpha* ABA thallus tissue demonstrated that the crude antiserum (200-fold diluted) reacted exclusively with a single polypeptide of the same size as the purified lectin (see Supplemental Fig. S9) indicating that the antibodies are monospecific. Accordingly, the antiserum was suitable for immunolocalization experiments without the need for further purification.

**Immunolocalization of ABA orthologs**

Samples of axenically cultured callus cells and small pieces of thallus were fixed with aldehyde (2% of formaldehyde and 0.25% of glutaraldehyde) in 0.05M cacodylate buffer (pH 7.2) for 24 h at 4°C, washed with cacodylate buffer and dehydrated in ethanol series (20, 40, 60, 75, 80, 95% and 100% of ethanol). After dehydration tissue samples were stepwise infiltrated with mixtures of LR White resin in ethanol (1/2, 1/1, and 2/1, respectively) and embedded in pure LR White resin. Polymerisation was finally performed overnight at 70°C. Transverse semi-thin (2 µm in thickness) were prepared using a Reichert ultraCutE microtome. Semi-thin sections were mounted on glass slides for bright field and laser scanning confocal microscopy examinations. Some sections were stained with toluidine blue for bright field microscopy examination. Sections were blocked for 2 h at room temperature in PBSTA (0.14 M NaCl, 2.7 mM KCl, 7.8 mM Na₂HPO₄, 1.5 mM KH₂PO₄, 1% de Tween 20, 1% BSA) (pH 7.2) and incubated overnight at 4°C with 10-fold diluted (in PBSTA) anti-Marpola antibodies. After washing with the same buffer sections were incubated for 2 h at room temperature with 1000-fold diluted (in PBSTA) anti-rabbit IgG coupled to Alexa 488 (Molecular Probes, Carlsbad CA, USA), washed and dried. Possible autofluorescence and unspecific binding of the
secondary labeled antibody were checked in a control experiment in which the primary anti-Marpola antibody was omitted.

For bright field microscopy, images toluidine blue stained sections were taken using a CCD camera (color Coolview, Photonic Science, Robertsbridge, UK). For immunolabeling, confocal images were acquired with a SP2 confocal laser scanning system equipped with an upright microscope (Leica, Germany) and a 40x (PL APO, N.A. 1.25) oil immersion objective. The 488 nm ray line of an argon laser was used to detect the labeling. Light emitted from the Alexa 488 probe was collected in the range between 500 and 540 nm. The laser line intensity and the PMT setting were kept constant for both the control and the labeled samples.

**Hemagglutination and hapten inhibition assays**

Agglutination assays were carried out in small glass tubes in a final volume of 50 µL containing 40 µL of a 1% suspension of red blood cells and 10 µL of extracts or lectin solutions. To determine the specific agglutination activity the lectin was serially diluted with two-fold increments. Agglutination was assessed visually after 1 h at room temperature. Erythrocytes were treated with trypsin as described previously (Van Damme et al., 1987).

The carbohydrate-binding specificity of the lectin was determined by hapten inhibition of the agglutination of trypsin-treated rabbit erythrocytes. To 10 µL of a solution of MarpoABA (10 µg mL⁻¹ in PBS) 10 µL aliquots of solutions of sugars (0.5 M in PBS) or glycoproteins (5 mg mL⁻¹ in PBS) was added. After preincubation for 1 h at 25°C, 30 µL of a 1% suspension of trypsin-treated rabbit erythrocytes was added and the agglutination evaluated after 1 h. To determine the inhibitory potency of the most active monosaccharides and glycoproteins the assays were repeated with serially diluted stock solutions of sugars and glycoproteins. The concentration required for 50% inhibition of the agglutination of trypsin-treated rabbit erythrocytes was determined visually.

**Glycan array screening**
A purified preparation of the lectin was labeled at a concentration of 2 mg mL\(^{-1}\) with an Alexa Fluor\textsuperscript{®} 488 Protein Labeling Kit (Molecular Probes, Carlsbad CA, USA) according to manufacturer’s protocol. The labeled lectin was applied to a PD-10 column (Sephadex G-25, Amersham Biosciences) and separated from free label. Protein concentration was determined by Lowry determination and Alexa 488 labeling efficiency determined at excitation 485 nm/emission 535 nm. The MarpoABA-Alexa 488 was applied to the Consortium streptavidin/biotin array and to the printed array for binding specificity determination (http://www.functionalglycomics.org/static/consortium/resources/resourcecoreh.shtml).

Labeled MarpoABA was screened on both arrays in binding buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM CaCl\(_2\), 2 mM MgCl\(_2\), 0.05% Tween 20, 1% BSA).

The lectin was screened on the streptavidin/biotin array (EA V3) as previously described (Bochner et al., 2005). Briefly, biotinylated glycosides (Korchagina and Bovin, 1992) were bound to streptavidin-coated microtiter plates in 4 replicates. Pre-coated plates were washed three times with 100 µL of wash buffer (binding buffer minus BSA) prior to incubation. A stock solution of MarpoABA-Alexa 488 (30 µg mL\(^{-1}\)) was added to each well of wash buffer on a Victor-2\textsuperscript{TM} 1420 Multi-label Counter (Perkin Elmer Life Sciences, Wellesley MA, USA) at excitation 485 nm/emission 535 nm.

The lectin was screened on the Consortium printed array (PA V1) as previously described (Blixt et al., 2005). Briefly, an aliquot of the same preparation of MarpoABA-Alexa 488 (300 µg mL\(^{-1}\)) was applied in a volume of 50 µL binding buffer onto the pre-printed slide, cover-slipped, and incubated protected from light for 1 hour. The slide was washed successively in wash buffer, wash buffer minus Tween 20, and deionized water, and then dried under a stream of nitrogen. The slide was read on a ScanArray Express Microarray Scanner and the image analyzed using ScanArray Express software (PerkinElmer Life Sciences).

Molecular modeling of ABA orthologs from Marchantia and Tortula
Hydrophobic cluster analysis, molecular modeling and docking experiments were carried out using standard techniques. Three-dimensional models of the proteins were built using the atomic coordinates of ABA (Carrizo et al., 2005). Full details are described in supplemental data.

**Phylogenetic analysis**

Multiple amino acid sequence alignments based on Clustal-X (Thompson et al., 1997) were carried out using SeqPup (Gilbert, D.G., Biology Dept., Indiana University, Bloomington, IN 47405, USA) and MacClade (Maddison and Maddison, 1992) was used to build a parsimony phylogenetic tree.

**Acknowledgements**

The authors are deeply indebted to Prof. Monaco for making available the X-ray coordinates of apoABA, ABA in complex with the T-antigen disaccharide and ABA in complex with the T-antigen disaccharide and GlcNAc, prior to release at the PDB. We also want to thank Dr Katsuyuki Yamato (Kyoto University, Japan) for providing the cDNA clone encoding MarpoABA2a (Genbank accession No. C95977) and Angela Lee of Core H of The Consortium for Functional Glycomics for expert technical assistance.

**Supplemental Data**

The following material is available from the website:

**Materials and methods:**

Molecular modeling of ABA orthologs from *Marchantia* and *Tortula*

**Figures S1 – S9:**
**Figure S1:** Alignment of the nucleotide sequences of the five contigs assembled from the available EST sequences of *Marchantia*

**Figure S2:** Alignment of the nucleotide sequences of the ORFs of the MarpoABA1 group (A) and their deduced amino acid sequences (B)

**Figure S3:** Alignment of the nucleotide sequences of the ORFs of the MarpoABA2 group (A) and their deduced amino acid sequences (B)

**Figure S4:** Alignment of the deduced amino acid sequences of the *T. ruralis* ABA ortholog with MarpoABA1a

**Figure S5:** Alignment of the nucleotide sequences of the ORFs of the five contigs assembled from the available EST sequences and an additional genomic sequence amplified by PCR

**Figure S6:** Mass spectrometry of *M. polymorpha* ABA orthologs

**Figure S7:** Analysis of carbohydrate binding specificity of *M. polymorpha* ABA orthologs using elisa glycan array and printed glycan array

**Figure S8:** Sequences of ABA orthologs from fungi and plants

**Figure S9:** Western blot analysis of ABA ortholog in crude extracts of *M. polymorpha* thalli

**Tables S1 – S2:**

**Table S1:** Overview of the main physicochemical parameters of the *M. polymorpha* and *T. ruralis* ABA orthologs
Table S2: Overview of the activity of the T-antigen dissacharide and GlcNAc-binding site of the M. polymorpha and T. ruralis ABA orthologs
Literature cited


Czapla TH, Lang BA (1990) Effect of plant lectins on the larval development of European corn borer (*Lepidoptera pyralidae*) and southern corn rootworm (*Coleoptera chrysomelidae*). J Econ Entomol 83: 2480-2485


Stewart CN Jr, Via LE (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14: 748-750


Figure legends

Figure 1. Structural alignment of the amino acid sequences of *M. polymorpha* ABA orthologs (MarpoABA), Tortula ABA ortholog (TorruABA) and ABA. Strands of β-sheet (β1-β10) and stretches of α-helix (α1, α2) occurring along the polypeptide chain of ABA have been indicated.

Figure 2. Localization of MarpoABA in thalli and callus of *M. polymorpha*. A to C: transverse sections of callus cells (A) and of upper (B) and lower (C) cell layers of the liverwort thallus after staining with toluidine blue. Note the rich content of parenchyma cells below the epidermal layer. Arrows (in C) indicate transverse sections of rhizoïds. D to F: confocal images of immunolabeled sections from the corresponding callus (D) and thallus (E,F) cells showing an immunolabeling essentially restricted to the cell contours. Scale bars = 20 µm.

Figure 3. Molecular modeling of ABA orthologs from *Marchantia* and *Tortula*. (A) Ribbon diagram of the ABA protomer showing the β-sandwich organization made of two bundles of β-sheet ([β1, β3, β4, β9, β10] and [β2, β5, β6, β7, β8]) interconnected by a helix-loop-helix consisting of two α-helices α1 and α2. N and C correspond to the N- and C-terminal ends of the polypeptide chain, respectively. The location of the T-antigen binding site (open star) and GlcNAc binding site (black star) are indicated by stars. (B) Ribbon diagram of the modeled MarpoABA2a. (C) Ribbon diagram of the modeled TorruABA. (D) Network of hydrogen bonds (dark dashed lines) connecting the T-antigen disaccharide (Gal β1,3GalNAc) (black sticks) to the amino acid residues forming the carbohydrate-binding site of ABA. (E) Docking of the T-antigen disaccharide (Gal β1,3GalNAc) (black sticks) in the carbohydrate-binding site of the modeled MarpoABA2a. (F) Network of hydrogen bonds (dark dashed lines) connecting GlcNAc (black sticks) to the amino acid residues forming the carbohydrate-binding site of ABA. (G) Docking of GlcNAc (black sticks) in the carbohydrate-binding site of the modeled MarpoABA1a.
Figure 4. Phylogenetic tree of ABA orthologs from fungi and plants. MarpoABA and TorruABA refer to the expressed plant proteins from *Marchantia polymorpha* and *Tortula ruralis*, respectively. Basidiomycota lectins are Agabi (*Agaricus bisporus* agglutinin, AAA85813); Paxin (*Paxillus involutus* expressed proteins AAT91249); Plecol (*Pleurotus cornucopiae* lectin PCL-F1, BAB63922); PlecoII (*Pleurotus cornucopiae* lectin PCL-F2, BAB63923); Phach1 (*Phanerochaete chrysosporium* hypothetical protein encoded by complement AADS0100313.1:6707-7153) and Xerch (*Xerocomus chrysenteron* lectin XCL1, AL73235). Ascomycota lectins are Artol (*Arthrobotrys oligospora* lectin, CAA65781); Gibze (*Gibberella zeae* hypothetical expressed protein EAA76455); Neucr (*Neurospora crassa* hypothetical expressed protein EAA30976); Podan (*Podospora anserina* hypothetical protein CAD60779) and Sorma (*Sordaria macrospora* hypothetical protein CAH03681). Figures refer to Genbank accession numbers. The scale bar indicates the number of amino acid changes.
Table I: Overview of the results of the binding assays of MarpoABA to the Consortium Glycan Arrays

<table>
<thead>
<tr>
<th>Glycans</th>
<th>PA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>EA&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Glycan&lt;sup&gt;b&lt;/sup&gt;</td>
<td>RFU&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Neu5Acα2-3Galβ1-3[6OSO3]GalNAcα-</td>
<td>163</td>
<td>40809</td>
</tr>
<tr>
<td>Neu5Acα2-3Galβ1-3GalNAcα-</td>
<td>166</td>
<td>39096</td>
</tr>
<tr>
<td>GlcNAcβ1-3Galβ1-3GalNAcα-</td>
<td>122</td>
<td>37850</td>
</tr>
<tr>
<td>Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα-</td>
<td>165</td>
<td>37378</td>
</tr>
<tr>
<td>Galβ1-3GalNAcα-</td>
<td>56</td>
<td>33681</td>
</tr>
<tr>
<td>β-D-Gal–</td>
<td>46</td>
<td>32572</td>
</tr>
<tr>
<td>α-GalNAc–</td>
<td>80</td>
<td>30639</td>
</tr>
<tr>
<td>Neu5Acβ1-6GalNAcα–</td>
<td>198</td>
<td>19026</td>
</tr>
<tr>
<td>6-Su-GalNAcα–</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6-Su-Neu5Acα2-3(Galβ1-3)GalNAcα-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup>PA and EA stand for printed array and elisa array, respectively.
<sup>b</sup>Glycan number refers to the number of the glycan in the respective arrays.
<sup>c</sup>Binding activity in PA and EA are expressed as RFU (relative fluorescence units) and S/N (signal over noise ratio), respectively.
<sup>d</sup>not present on the array.