Biochemical characterization of the novel endo-β-mannanase AtMan5-2 from Arabidopsis thaliana

Yang Wang, Shoaib Azhar, Rosaria Gandini, Christina Divne, Ines Ezcurra, Henrik Aspeborg

Abstract

Plant mannanases are enzymes that carry out fundamentally important functions in cell wall metabolism during plant growth and development by digesting manno-polysaccharides. In this work, the Arabidopsis mannanase 5-2 (AtMan5-2) from a previously uncharacterized subclade of glycoside hydrolase family 5 subfamily 7 (GH5,7) has been heterologously produced in Pichia pastoris. Purified recombinant AtMan5-2 is a glycosylated protein with an apparent molecular mass of 50 kDa, a pH optimum of 5.5–6.0 and a temperature optimum of 25 °C. The enzyme exhibits high substrate affinity and catalytic efficiency on mannose substrates with main chains containing both glucose and mannose units such as konjac glucomannan and spruce galactoglucomannan. Product analysis of manno-oligosaccharide hydrolysis shows that AtMan5-2 requires at least six substrate-binding subsites. No transglycosylation activity for the recombinant enzyme was detected in the present study. Our results demonstrate diversification of catalytic function among members in the Arabidopsis GH5,7 subfamily.

1. Introduction

Mannan polysaccharides are non-crystalline hemicelluloses that exist in various organisms such as bacteria, fungi and plants. Higher-plant mannanas are involved in several biological processes during cell wall degradation and modification [1,2], but also serve as raw biopolymers that are widely used in food industry, paper pulping, and biofuel production [3–5]. As structural cell-wall components, mannan polysaccharides are suggested to cross-link or coat cellulose microfibrils, and to form linkages to other hemicelluloses and lignin, although mannan may also serve as carbohydrate reserves in seeds and tubers [6]. All mannan types contain a backbone of β-1,4-linked mannose units, whereas β-1,4-linked mannosyl and glucosyl residues can presumably be randomly distributed in the backbones of heterogeneous mannans [7]. Additionally, mannose residues in the main chain can be grafted by α-1,6-linked galactose units forming side chains. Based on the main-chain composition and the presence or absence of galactose side chains, mannan polysaccharides are typically classified into four major groups: pure mannans, glucomannans, galactomannans, and galactoglucomannans. In the model plant Arabidopsis, mannan polysaccharides are present as minor components in most cell types [8–11]. Glucomannan is the mannan polysaccharide found in the Arabidopsis stem and the seed mucilage [8,12], but whether other types of mannans exist in Arabidopsis tissues is currently not known [8,9,13]. Galactomannans are commonly found in seed endosperms from various plants but have not yet been found in Arabidopsis endosperm tissue [14].

Mannan-active enzymes play a key role in the natural recycling of biomass, and in the growth and development of plants [15]. They also serve as valuable biotechnological tools in industrial applications that utilize mannans [16,17]. Endo-acting β-mannanases (MANs) cleave the internal glycosidic bonds of the β-1,4-linked backbone of various mannans, and they constitute the principal enzymes involved in mannan depolymerization and/or modification. Based on protein sequence similarity, MANs characterized to date belong to either of the glycoside hydrolase (GH) families 5, 26 or 113, as classified according to the CAZy (http://www.cazy.org) database of carbohydrate-active enzymes [18]. Enzyme members of these GH families share a common three-dimensional (3D) fold known as the (β/α)5-barrel, or TIM barrel fold, and a similar catalytic machinery that hydrolyzes glycosidic bonds by retention of
configuration at the anomeric carbon, i.e., a double-displacement ($S_N$1) mechanism [19,20]. An inherent ability of many retaining glycoside hydrolases is to catalyze the “reverse” reaction whereby a sugar moiety is used as an acceptor rather than a water molecule during the deglycosylation step (the second step of the retaining reaction) leading to the formation of a new glycosidic bond, i.e., transglycosylation. Indeed, many microbial MANs and several plant MANs have been reported to exhibit both endo-hydrolytic and transglycosylation activity [21,23–26].

All sequenced genomes of higher plants contain genes coding for GH5 subfamily 7 (GH5,7) MANs [25,27] and a phylogenetic analysis of plant GH5,7 members revealed two principal clades of MANs [25]. Enzymatically characterized enzymes from tomato (Solanum lycopersicum, previously Lycopersicon esculentum), soybean (Glycine max), thale cress (Arabidopsis thaliana), barley (Hordeum vulgare) and coffee (Coffea arabica) are represented in clade 2 [21,24,25,28,29], whereas the only characterized enzyme in clade 1 is the Populus trichocarpa PtmAN6 [30]. It is not clear why plants need multiple forms of MANs, but one reason may be that the isoenzymes have evolved to exhibit different activities towards the various mannan substrates existing in plants.

To investigate the extent of catalytic diversification of plant GH5,7 enzymes, Arabidopsis provides an excellent plant model since its genome features seven full-length genes coding for MANs. Thus far, only a single characterized Arabidopsis MAN has been reported, namely the A. thaliana mannanase 1 (AtMan5-1). AtMan5-1 belongs to clade 2, and was shown to hydrolyze carob galactomannan, konjac glucomannan, spruce galactoglucomannan (sGGM) and mannopentaose (M5), as well as generating transglycosylation products [25]. Another Arabidopsis MAN, A. thaliana mannanase 2 (AtMan5-2; locus At2g20680), belongs to clade 1 of the GH5,7 phylogenetic tree, and seed germination transcript
profiling has suggested a role for the enzyme in this process [31]. However, in contrast to the other MAN genes expressed in germinating seeds (AtMan5-5, AtMan5-6 and AtMan5-7), the T-DNA insertion mutant of AtMan5-2 did not affect the germination time course, indicating that the function of AtMan5-2 is different from that of the other MANs associated with seed development [31].

In order to shed light on the function and catalytic properties of AtMan5-2, we have produced heterologously AtMan5-2 as a functional enzyme in *Pichia pastoris*. AtMan5-2 is the first enzyme to be characterized in this subclade of the GH5,7 clade 1 enzymes, which are here shown to feature an unusual +1 substrate binding subsite and two conserved cysteine (Cys) residues in the C-terminal region. When comparing the catalytic properties of AtMan5-2 to those of the previously reported AtMan5-1, we observe distinct differences in the substrate specificity between the two enzymes representing the two major evolutionary branches of plant GH5,7 MANs. This work reports, for the first time, on the catalytic diversification of plant MAN paralogs.

2. Materials and methods

2.1. Cloning, recombinant protein production and purification

To adopt the nomenclature of Arabidopsis GH5,7 MANs, the enzyme encoded by the gene identified as At2g20680 (AtMAN2) was named AtMan5-2 [25]. The plasmid containing the full-length cDNA sequence encoding AtMan5-2 was obtained from RIKEN [32,33]. A gene construct containing the catalytic domain (K35 to P433) of AtMan5-2 without the putative signal peptide, was amplified by PCR using Phusion™ High-Fidelity Polymerase (Finzymes) and optimized primers (forward primer: ATTATTGGCCGCGAAAACGGAGGCCGAG; reverse primer: GGCG-GCTCTAGATGTTGCTATGGAAACAC). The gene was cloned into the pPICZα C vector (Life Technologies) to generate a construct including the *Saccharomyces cerevisiae* α-factor secretion signal for secreted expression, and a C-terminal His6-tag for convenient protein purification by immobilized metal affinity chromatography (IMAC). The identity of the target gene was verified by DNA sequencing (Eurofins MWG Operon). To produce the secreted target protein, *P. pastoris* SMD1168H transformed with the recombinant plasmid was grown as described for AtMan5-1 in our previous work [25].

Recombinant His6-tagged AtMan5-2 was purified by IMAC using a HiTrap IMAC FF column (GE Healthcare) equilibrated with 20 mM sodium phosphate, 0.5 M sodium chloride, 5 mM imidazole, pH 7.4 (buffer A) mounted on an ÄKTA purifier system (GE Healthcare Life Sciences). The culture supernatant was loaded onto the column and washed with buffer A, after which linear gradient elution of 5–500 mM imidazole was performed in 20 mM sodium phosphate buffer including 0.5 M sodium chloride (pH 7.4). The purity and molecular mass of the recovered protein was evaluated using SDS-PAGE analysis (12% precast polyacrylamide gels, BIO-RAD, U.S.A.). The protein was incubated in the presence of reducing agents and heated at 95 °C for 5 min before loading on a gel. Protein concentration was determined by Bradford assay [34] using the BioRad Protein Assay Kit (Sweden). To confirm the identity of the purified protein, a single protein band with the expected size was cut out from the gel, digested by trypsin and identified by mass spectrometry (Science for Life Laboratory, Sweden).

2.2. Deglycosylation and analysis of the oligomeric state

Deglycosylation of recombinant AtMan5-2 was performed by treating the protein with Endoglycosidase H₁ (Endo H₁; New Eng- land BioLabs) following the manufacturer’s instructions, and the result was analyzed by SDS-PAGE. To investigate whether reducing agents affect the oligomeric state of the AtMan5-2, the purified protein was incubated in the presence or absence of 50 mM dithiothreitol (DTT), and heated at 75 °C for 7 min, followed by SDS-PAGE analysis.

2.3. Biochemical characterization

The hydrolytic activity of AtMan5-2 was determined by quantifying the amount of reducing sugar released using the 3,5-dinitrosalicylic acid (DNS) reducing sugar assay [35], as described earlier [25]. One unit of MAN activity was defined as the amount of enzyme required to release reducing sugar equivalents to 1 μmol of mannose per min under the given conditions. The assay mixtures contained mannan poly- or oligosaccharides (all except sGGM were from Megazyme) and 0.2–1.9 μM AtMan5-2 in 50 mM sodium acetate buffer. To determine the temperature profiles, thermostability of activity as well as effect of glycosylation and metal ions on AtMan5-2 activity, activity assays were performed using 3 mg/mL konjac glucomannan in 50 mM sodium acetate buffer (pH 5.5) for 30 min.

The pH profile was determined using different 50 mM buffers: sodium formate pH 3.0–4.5; sodium acetate, pH 4.5–5.5; sodium citrate, pH 5.5–6.0; and sodium phosphate, pH 6.0–7.0. The temperature profile was determined within the temperature range 0–45 °C in steps of 5 °C. To determine thermostability of activity, aliquots of enzyme (1.5 μM) were incubated at different temperatures in the range 15–45 °C in steps of 10 °C. Following incubation, the residual activity was measured at different time points, i.e., 30, 60, 120 and 240 min.

To assess the effects of freeze–thawing, enzyme samples were stored at ~20 °C, followed by measuring the residual activity at different time points, i.e., 1, 2, 4 and 22 h. To analyze the effect on activity of treating AtMan5-2 with Endo H₁ to remove surface-attached N-glycans, the enzyme was incubated in the absence (control) or presence of Endo H₁ at 37 °C for 2 h, immediately followed by activity measurements.

To analyze the effect of reducing agents on the activity, the enzyme was incubated in the presence of reducing agent, i.e., 1, 10 and 50 mM DTT/β-mercaptoethanol (βME) in 20 mM sodium phosphate buffer (pH 7.4) at room temperature for 1 h. Following incubation with reducing agent, the residual activity was determined using a reaction mixture containing 4 mg/mL konjac glucomannan at the given concentration of reducing agent at optimal reaction conditions for 4 h, i.e., sodium acetate buffer (50 mM, pH 5.5) and 25 °C. Protein samples without reducing agent were used as controls and treated identically to those containing reducing agent.

The influence of metal ions on AtMan5-2 activity was analyzed for a range of chloride salts, i.e., Ca²⁺, Co²⁺, Fe³⁺, Li⁺, Mg²⁺ and Ni²⁺. AtMan5-2 was incubated with 5 mM EDTA and EGTA at room temperature for 1 h, and washed with 50 mM sodium acetate buffer (pH 5.5) using 30 kDa cutoff Amicon Ultra centrifugal filters (Millipore). Then, metal salts were added to the enzyme solutions to a final concentration of 5 mM, and the activity measured after 30 min incubation. Blank controls were prepared in the same way as the samples but using the 20 mM sodium phosphate buffer instead of enzyme. To determine and compare the activity of AtMan5-2 on different mannan-polysaccharides; carob galactomannan (1.0–5.0 mg/mL), konjac glucomannan (1.5–6.0 mg/mL) and sGGM (2.0–6.0 mg/mL) were incubated with enzymes in sodium acetate buffer (50 mM, pH 5.5) at 25 °C. Due to the polydisperse nature of the mannan substrates used in the reactions, the apparent Michaelis constant *Kₘ* was expressed in mg/mL. The apparent *Kₛ* was calculated as the maximum catalytic formation
of product (µmol) per unit enzyme (µmol) per unit time (s) under saturating substrate condition.

2.4. Product analysis from mannan-polysaccharides and oligosaccharides

To analyze the hydrolysis products, AtMan5-2 was incubated with 1 mg/mL mannan polysaccharides (carob galactomannan, konjac glucomannan, guar gum, SGM and carboxymethyl cellulose) or 1 mM manno-oligosaccharides (mannotriose, M3; mannotetraose, M4; mannopentaose, M5; mannohexaose, M6; and cellohexaose, C6) at 25 °C for 48 h. The transglycosylation activity of AtMan5-2 was investigated under conditions that favor transglycosylation, i.e., high concentrations of substrate and salt. For this purpose, enzyme samples (1.9 µM) were incubated with 5 mM M5/M6 with or without 1 M NaCl at 25 °C for 24 h. Hydrolysis products were identified and analyzed by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) as reported previously [25].

2.5. Protein sequence analysis, homology modeling and ligand docking

Protein sequences used for bioinformatics analyses were retrieved from Phytozome version 10 (http://phytozome.jgi.doe.gov). Multiple sequence alignment was performed using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2). To analyze conserved residues in AtMan5-2 and the related plant GH5,7 enzymes in the AtMan5-2 subclade, a multiple sequence alignment was performed using 38 protein sequences, and the data presented by WebLogo analysis (http://weblogo.threeplusone.com/create.cgi). Sequence motifs for N-linked glycosylation were predicted by NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/), and phosphorylation sites were predicted by NetPhos 2.0 Server (http://www.cbs.dtu.dk/services/NetPhos/).

Homology models of AtMan5-2 were generated automatically using RaptorX [36], PHYRE2 [37] and SWISS-MODEL [38]. The geometry of the homology models was analyzed using MolProbity (http://molprobity.biochem.duke.edu/; [39]). All models had acceptable geometry, but the model generated by SWISS-MODEL performed best with respect to coarse and fine packing quality as determined by WHAT IF [40]. To analyze the possible sugar-binding subsites in AtMan5-2, an extended manno-oligosaccharide was docked manually to the binding cleft of the SWISS-MODEL homology model. Modeling of the oligosaccharide was guided by sugar binding in the crystal structures of: Trichoderma reesei (anamorph of Hypocreajecorina) TrMan5A in complex with M2 bound to subsites +1 and +2 (PDB code 1QNR; [41]); Rhizomucor miehei RmMan5B E202A variant in complex with mannotriose (M3) bound to subsites –1, +1 and +2 (PDB code 4LYQ; [42]); and Thermomonospora fusca TjMan in complex with M3 bound in subsites –4, –3 and –2 (PDB code 3MAN; [43]). More precisely, the three crystal structures and the

Fig. 2. Optimum activity of AtMan5-2 using konjac glucomannan as substrate as a function of pH (a), and temperature (b). For calculation of residual activity, the optimum activity was set as 100%. (c) Thermostability of AtMan5-2 activity using konjac glucomannan as substrate as a function of temperature and time: 15 °C (black squares); 25 °C (red circles); 35 °C (blue triangles); and 45 °C (pink inverted triangles). For calculation of residual activity, the activity at t = 0 was set as 100%. (d) Freeze-thaw resistance of AtMan5-2 activity. For calculation of residual activity, the activity without freeze thawing was set as 100% activity. Time is the freezer storage time.
theoretical AtMan5-2 model were superimposed and the sugars bound to the individual subsites in the experimental crystal structures were adjusted manually to optimize the fit of the sugars in the predicted subsites of the AtMan5-2 model. Based on the structure of RmMan5B with bound M3 [42], the mannose in subsite –1 was modeled as a 155 skew boat in the AtMan5-2 model.

2.6. Analysis of expression and upstream sequences

The expression of the AtMan5-2, AtMan5-5, and POPTR_0013s13400 genes was investigated using the public microarray database BAR eFP [44,45]. Upstream promoter sequences (1500 bp from ATG) of AtMan5-2 orthologs were retrieved from Phytozome. De novo motif analysis was performed by using MEME (http://meme.ewi.edu.au/meme/index.html [46]).

3. Results and discussion

3.1. Protein production and purification

The catalytic domain of AtMan5-2 carrying a C-terminal His6-tag was successfully produced in P. pastoris SMD1168H cells, and purified to homogeneity with a yield after IMAC of 16 mg/L. The recombinant fusion protein including the AtMan5-2 catalytic domain, consisting of 399 amino-acid residues, a c-myc epitope, a His6-tag and an additional peptide sequence in between the signal sequence processing sites and the start of the catalytic domain migrates on SDS-PAGE gels as a distinct band with an approximate molecular mass of 50 kDa (Fig. 1a). Mass spectrometry (MS) analysis identified 30 unique peptides that matched the amino acid sequence of AtMan5-2 (Supplemental Fig. S1). The MS analysis also revealed phosphorylation of Ser81, although only approximately 8% of the peptides carrying Ser81 were identified as phosphopeptides. Phosphorylation of AtMan5-2 in Pichia indicates that the protein has potential to be phosphorylated also in planta. Treatment with Endo Hf to remove N-linked glycans resulted in a decrease in molecular weight (Fig. 1b), thus confirming that recombinant AtMan5-2 is generated by P. pastoris as a glycoprotein. The apparent molecular weight of the deglycosylated AtMan5-2 is estimated to be around 5 kDa lower than the calculated molecular weight of approximately 50.7 kDa.

The first 28 amino acids were not included in the gene construct since these are predicted to constitute a signal peptide. Furthermore, the C-terminus of the produced protein appears to be intact since the C-terminal His6-tag was positively identified using a His probe and Western Blot analysis (data not shown). The low coverage for the C-terminal region by MS analysis is therefore not due to prematurely terminated translation or proteolytic cleavage. Thus, the aberrant migration of AtMan5-2 on SDS-PAGE cannot be explained by truncation, but rather, may be due to factors such as protein sequence composition, posttranslational modifications and/or structural features that may influence protein electrophoresis migration.
No support for glycosylation was obtained from the MS analysis, because the peptides containing the two predicted glycosylation sites were not identified. Glycosylation was also reported for the *Pichia*-produced AtMan5-1 enzyme, although the recombinant AtMan5-1 appears to be more extensively glycosylated compared with the heterologously produced AtMan5-2 [25]. Glycosylation of recombinant proteins produced in heterologous eukaryotic hosts will often differ from the native protein. Consequently, we do not know if N-glycans are attached to the AtMan5-2 and AtMan5-1 proteins *in planta*, but glycosylation of a poplar MAN has been observed [30], so it seems reasonable to expect that these Arabidopsis enzymes are native glycoconjugates. Under non-reducing conditions, AtMan5-2 migrates as two bands on the SDS-PAGE gel (Fig. 1c, lane 2), one major band corresponding to the 50 kDa monomer and one minor band with the apparent mass of a dimer. Under reducing conditions, however, only the monomeric band is present (Fig. 1c, lane 3) suggesting that intermolecular disulfide bonds may mediate dimer formation. To further investigate possible AtMan5-2 oligomerization, size exclusion chromatography (SEC) analysis was performed, and using this method only the monomer was observed in solution (Supplemental Fig. S2). There is a slight leading shoulder in the gel-filtration profile of AtMan5-2 that may indicate traces of a dimeric species. Thus, AtMan5-2 can exist as a dimer, but the predominant form in solution is the monomer.

3.2. Determination of pH and temperature dependence of activity and thermostability of activity

The hydrolytic activity of AtMan5-2 was studied as a function of temperature and pH using konjac glucomannan as substrate. AtMan5-2 displayed maximum activity at pH 5.5–6.0 with an apparent optimal temperature at 25°C (Fig. 2a,b). The activity was also examined as a function of temperature and time (Fig. 2c). After four hours of incubation at temperatures up to 35°C, AtMan5-2 retained 90–100% of the initial activity. At 45°C, however, the residual activity remaining after 4 h was 45%. Freeze-thawing was also found to affect AtMan5-2 activity resulting in only 20% residual activity after 22 h incubation at −20°C (Fig. 2d). The optimal pH for AtMan5-2 activity is similar to the apparent pH optima reported for the *Pichia*-produced AtMan5-1 [25] and other characterized plant MANs [24,28]. The two Arabidopsis GH5 MANs also show similar thermostability of activity, although the optimal temperature for activity was 10°C lower for AtMan5-2 compared with AtMan5-1.

3.3. Influence of glycosylation, reducing agents and metal ions on activity

Next we investigated whether the observed glycosylation could affect catalysis. The influence of glycosylation on AtMan5-2 activity was assessed by removing N-linked glycans using Endo H. No significant decrease in activity was observed after deglycosylation at 37°C for 2 h (Fig. 3a), and this is in contrast to *P. manihot* where N-glycosylation was reported to be essential for catalytic activity [47]. Nevertheless, recombinant LeMAN4A from tomato and AtMAN5-1 were catalytically active although generated in *Escherichia coli* [24,25], so plant MAN activities seem to be diversely affected by protein glycosylation. Contrary to deglycosylation, addition of the reducing agents DTT and βMe significantly reduced AtMan5-2 activity (Fig. 3b,c). Treatment with 50 mM DTT and βMe caused relative activity decreases of approximately 80% and 60%, respectively. Effects of various metal ions known to affect MAN activity were tested and these chemicals affected AtMan5-2 activity to varying degree (Fig. 3d). While Mg²⁺ enhanced the relative activity, Ni²⁺ and Fe³⁺ reduced the activity to some extent, other cations showed marginal effects on activity. The precise mechanism of action for these divalent cations is not clear, but most likely they play a role by stabilizing local structure critical for activity as seen for an actinomycetes MAN [46].

3.4. Determination of kinetic constants for manno-poly saccharides

The kinetic parameters were determined for the natural mannan polysaccharides: carob galactomannan, konjac glucomannan and sGGM (Table 1). Amongst the substrates that were tested, konjac glucomannan displayed the highest Vₘₐₓ (6.7 U/mg), whereas lower maximal velocities were obtained with carob galactomannan (2.5 U/mg) and sGGM (1.7 U/mg). sGGM showed the lowest Michaelis constant (sGGM, Kₘ = 0.5 mg/mL; konjac glucomannan, Kₘ = 1.3 mg/mL; carob galactomannan, Kₘ ~ 12.8 mg/mL). The activity on guar gum galactomannan was too low to be accurately measured. When comparing the obtained catalytic properties of AtMan5-2 with those determined for AtMan5-1 [25] some differences are noted. While AtMan5-2 seems to have a preference for substrates with a mannan backbone containing glucose, AtMan5-1 is able to hydrolyze carob galactomannan, konjac glucomannan and sGGM at comparable efficiencies. These results suggest glucomannan may be the *in vivo* substrate for AtMan5-2.

3.5. Product patterns from manno-polysaccharides and oligosaccharide hydrolysis

To further elucidate the AtMan5-2 enzymatic mode of action, product patterns of hydrolysis using manno-polysaccharides and oligosaccharides were analyzed using HPAE-PAD. The major mannan oligosaccharide products from carob galactomannan hydrolysis were mannobiose (M2) and M4, as well as a smaller peak identified as M5 (Fig. 4a). Hydrolysis of konjac glucomannan generated mainly M2 and M4 with minor amounts of M3 and M5, and for sGGM, weak product traces of M2, M4 and M5 were detected. With guar gum galactomannan as substrate, only a weak trace of M4 was produced. Both AtMan5-2 and AtMan5-1 showed negligible activity toward the highly branched guar gum galactomannan, but the hydrolytic cleavage profiles obtained for the other substrates were divergent. For instance, AtMan5-1 produced M2, M3 and significant amounts of GM3 from carob galactomannan, but M2 and M3 from konjac glucomannan with M2 being the principal product [25]. Unidentified and overlapping peaks observed in the experiments probably represent short glucomannno-oligosaccharides or branched galactomannano-oligosaccharides.

Furthermore, the present study revealed that AtMan5-2 was not able to degrade carboxymethyl cellulose (CMC), nor C6, M3, M4 and M5 (data not shown), whereas M6 was hydrolyzed to give M2, M3 and M4, with M3 as the predominant species (Fig. 4b). The absence of reaction products released from M3-M5 incubated with the AtMan5-2 revealed the requirement of at least six substrate-binding subsites. Although the hydrolysis product patterns of M6 are similar for AtMan5-2 and AtMan5-1, the latter enzyme was also able to degrade M5 generating M2 and M3. This suggests that both enzymes have at least six sugar-binding subsites. The affinity of the individual subsites differ to generate catalytic diversification between these two Arabidopsis GH5 MANs.

Since *in vitro* transglycosylation activity has been reported for several plant MANs [21,24,25], attempts were made to find evidence for transglycosylation. The absence of manno-oligosaccharides of higher degree of polymerization (DP) than M6, was interpreted as no significant transglycosylation had occurred even at high substrate concentration or with the addition of salt [24]. However, any transient transglycosylation products hydrolyzed by AtMan5-2 may not have been measurable by the method. On the other hand, transglycosylation products were
Table 1

<table>
<thead>
<tr>
<th>Manno-polysaccharides</th>
<th>( V_{max} ) (U/mg)</th>
<th>( K_m ) (mg/mL)</th>
<th>( k_{cat} ) (s(^{-1}))</th>
<th>( k_{cat}/K_m ) (s(^{-1}) mg(^{-1}) mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carob galactomannan</td>
<td>2.5 ± 0.08</td>
<td>12.8 ± 0.38</td>
<td>2.1 ± 0.06</td>
<td>0.2 ± 0.0001</td>
</tr>
<tr>
<td>Konjac glucomannan</td>
<td>6.7 ± 0.14</td>
<td>1.3 ± 0.04</td>
<td>5.7 ± 0.12</td>
<td>4.3 ± 0.05</td>
</tr>
<tr>
<td>Spruce galactoglucomannan</td>
<td>1.7 ± 0.06</td>
<td>0.5 ± 0.04</td>
<td>1.5 ± 0.05</td>
<td>2.8 ± 0.12</td>
</tr>
</tbody>
</table>

Fig. 4. (a) HPAEC-PAD oligosaccharide profiles following enzymatic treatment of different mannan poly- and oligosaccharides. (b) HPAEC-PAD profiles of mannoheaxose (M6) and mannoheaxose with increased substrate concentration and addition of salt (M6 + S). Mannopentaose (M5) was present as a contaminant in the M6 sample as shown in controls without enzyme (Blank M6) and (Blank M6 + S). M2, mannobiose; M3, mannotriose; M4, mannotetraose; M5, mannopentaose; M6, mannoheaxose.

detected for recombinant AtMan5-1 incubated with M5 and M6 under similar conditions favoring transglycosylation as used in the present study. The observed contrasting behavior of AtMan5-2 and AtMan5-1 likely illustrates different capacity to perform transglycosylation reactions and is yet another sign that these enzymes have evolved to function differently.

3.6. Protein sequence analysis

The AtMan5-2 gene codes for 433 amino-acid residues with a putative N-terminal signal peptide (residues 1–28) for secretion. We have reported earlier that AtMan5-2 and AtMan5-5 belong phylogenetically to the same subclade of clade 1, one of the two major clades of plant GH5,7 MANS [25]. This implies that these proteins are closely related. A protein sequence alignment including the clade-1 MANs (AtMan5-2, AtMan5-5, and PtrMAN6) and biochemically characterized clade-2 MANs (AtMan5-1 and LeMAN4a) is shown in Fig. 5a. The sequences of AtMan5-2 and AtMan5-5 are indeed similar with 83% sequence identity.

Of the enzymes in clade 1, only PtrMAN6 has been characterized biochemically, but this enzyme belongs to a different subclade of clade 1 [30,47]. The two glutamate residues performing the catalytic roles as acid/base (Glu215 in AtMan5-2) and nucleophile (Glu335 in AtMan5-2) are conserved in all aligned sequences (Fig. 5a). Other residues that are typically conserved among GH5 enzymes are found also in AtMan5-2, i.e., Arg96, His293, Tyr295 and Trp377. Plant MANs have been reported to be generated as glycoproteins [25,30], and in this study AtMan5-2 was produced as a glycosylated recombinant protein by P. pastoris. There are two N-X-T/S sequence motifs for N-glycosylation in AtMan5-2, 46NGT48 and 169NDS171, of which the former exists in AtMan5-2, AtMan5-1 and AtMan5-5, and the latter only in AtMan5-2 and AtMan5-5 (Fig. 5a). The Ser81 identified as being phosphorylated with MS was predicted as a potential phosphorylation site using NetPhos 2.0, although this software has not been trained using secreted phosphoproteins.

The fact that thiol-reducing agents affected the oligomeric state of AtMan5-2 motivated a search for Cys residues in the protein sequences. PtruMAN6 and its plant orthologs feature three conserved Cys residues in the C-terminal region: Cys448, Cys452 and Cys456 [47]. These Cys residues have been confirmed experimentally to play a role in dimerization of PtruMAN6, and also to affect the enzyme’s catalytic activity [30,47]. One of the Cys residues (Cys456 in PtruMAN6) is present in AtMan5-2 (Cys429) and AtMan5-
Fig. 5. (a) Protein sequence alignment including AtMan5-2 (locus At2g02680, UniProtKB accession Q7Y223), and selected GH5_7 plant MANs from A. thaliana (AtMan5-1, locus At1g02310, UniProtKB accession Q9FF29; AtMan5-5, locus At4g28320, UniProtKB accession Q9M0H6), S. lycopersicum (LeMAN4a, UniProtKB Q8L5J1) and P. trichocarpa (PtrMAN6, UniProtKB accession B8ICV3). Two predicted glycosylation sites (46NGT48 and 169NDS171) are shown and two catalytic glutamic acid residues are framed by squares. Arrows denote additional amino acids in GH5 members that are typically conserved. The arginine positioned at substrate binding site +2 is marked with a circle. (b) WebLogo analysis of the C-terminal region of a sequence alignment using the amino acid sequences of AtMan5-2 and 38 GH5_7 plant MANs belonging to the same subgroup as AtMan5-2. Arrows highlight two highly conserved Cys residues.
5 (Cys427). We also identified a second Cys residue present in AtMan5-2 (Cys410) and other members of this particular subclade (Fig. 5b). Possibly, these Cys residues have a functional role in AtMan5-2 and closely related plant MANs. Interestingly, Pichia-produced recombinant AtMan5-2 protein can also exist as dimers in vitro under non-reducing conditions, leading to a polydisperse protein solution consisting of monomers (main species) and dimers (minor species). Thus, it is conceivable that at least one of the Cys found in the C-terminal part may play a role in disulfide-mediated dimerization of AtMan5-2; however, the function of dimerization is not clear. Although the enzyme activity can be inhibited by DTT and βMe, it is possible that these compounds interact directly with the active site to cause inhibition since the AtMan5-2 monomer is the predominant species.

The function of the N-terminus in AtMan5-2 is not known but may include a secretion signal. To further analyze the likelihood of a signal peptide at the N-terminus, the AtMan5-2 sequence was analyzed using SignalP 4.1 [49], Phobius [50], and TOPCONS [51]. Interestingly, SignalP 4.1 fails to identify a signal peptide in AtMan5-2 but reports a possible transmembrane helix, while both Phobius and TOPCONS report a strong signal for a transmembrane helix for residues 12–34 and 8–28, respectively. Hence, it is possible that AtMan5-2 is a membrane-anchored enzyme, rather than secreted to the Arabidopsis cell wall.

3.7. Structural analysis of the theoretical homology models

Several homology models of AtMan5-2 were generated. The closest homolog with known 3D structure is LeMAN4a (PDB code 1RH9; [52]; 43% sequence identity). The LeMAN4a template ends at Ser399 (Thr416 in AtMan5-2) and therefore no trustworthy model could be generated for the C-terminal region where Cys410 and Cys429 are located. In many GHs, the exo versus endo modes of action is affected by the length of active-site loops [42,53]. This is true for exo-acting CmMan5A and RmMan5B where a long loop (378–412 in CmMan5A; 354–392 in RmMan5B) forms a double-steric block to exclude glycone subsites beyond subsite −1 [42]. In our model of endo-acting AtMan5-2, and LeMAN4a, this loop is shorter (380–394 in AtMan5-2; 363–377 in LeMAN4a), which creates an extended cleft that runs across one face of the molecule (Fig. 6a,b). Furthermore, a manno-oligosaccharide (M7) was docked in the AtMan5-2 subsite +3 (aglycone side) to subsite −4 (glycone side) of the cleft that covers a length of 40 Å (Fig. 6c,d). With hydrolysis taking place between subsites −1 and +1, M7 is expected to generate mainly M3 and M4. The inability of AtMan5-2 to hydrolyze manno-oligosaccharides shorter than M6, and M3 being the principal hydrolysis product from M6, is in agreement with the modeled AtMan5-2 complex. Although mainly M3 is produced from M6, there are small amounts of M2 and M4 generated (Fig. 4b). This can be interpreted as four “negative” subsites (−4 to −1) and two of the possible “positive” subsites (+1 and +2) being used, thus implying four “negative” subsites rather than three, which is also supported by docking.

The “roof” of the cleft displays high sequence conservation with LeMAN4a whereas the “floor” is less well conserved, including the far aglycone end at subsite +3 where Arg269 and Trp270 lack counterparts in LeMAN4a (Fig. 6e). Tyr105 and Phe387 may interact with a mannone in subsite −4, while Asp102 could interact in subsite −3 (Fig. 6f). In subsite −2, Asp389, Trp98, and possibly also Gln146 may provide interactions. In subsite −1, a distorted mannone may interact with Asn214 and Trp377 (Fig. 6g). The acid/base catalyst Glu215 and nucleophile Glu335 would be suitably positioned for bond cleavage at subsites +1/−1. In subsite +1, a modeled mannone unit may hydrogen bond to Glu252, while Leu145 could offer hydrophobic interactions (Fig. 6h). Trp299 may interact in subsite +2, while Arg269 and Trp270 are positioned in subsite +3 (Fig. 6i).

GH5 MANs can accommodate both mannose and glucose in subsites −2 and +1 [54]. In our model, an axial O2 hydroxyl group appears favorable in subsites −2 and −1, which may suggest that these are mannose-recognition subsites rather than binding sites for glucose.

A tryptophan in subsite +1 in Aspergillus niger ManBK (Trp112) has been studied by mutagenesis [55]. The substitution Trp112→Gly resulted in a 13-fold increase in $k_m$ and 26-fold decrease in $k_{cat}$, whereas for Trp112→Tyr the $k_m$ value was unchanged and $k_{cat}$ reduced 4.5-fold. This suggested that an aromatic residue is preferred in subsite +1. However, AtMan5-2 subclade members have a leucine in this position (Leu145; Fig. 6h). The precise meaning of this discrepancy is unclear and needs further investigation by site-directed mutagenesis; however, hydrophobic interaction of a sugar with Leu is probably less favorable than a sugar-aromatic stacking interaction with the indole ring of a Trp.

Interactions in the aglycone subsites are thought to influence transglycosylation [22,56,57], especially an arginine in subsite +2 of fungal MANs [56]. In TrMan5A, Arg171 in subsite +2 offers two hydrogen bonds to O2 and O3 of a mannone residue, as well as two ionic interactions with Glu205 (Glu252 in AtMan5-2) [41]. The TrMan5A variant_R171K showed reduced transglycosylation and altered transglycosylation-product pattern [56]. Although this Arg is present also in AtMan5-2 (Arg217), no transglycosylation was observed. Probably, transglycosylation is more complex and not due to any single amino acid. Conceivably, weaker binding in subsite +1 due to Leu145 (Fig. 6h) may at least partly account for the loss of transglycosylation activity for AtMan5-2.

3.8. Analysis of gene expression and conserved cis-regulatory motifs

Expression analyses by RT-PCR indicate that AtMan5-2 is expressed in stem, roots, rosette leaves and germinating seeds, whereas AtMan5-5 is weakly expressed in stem [58]. Moreover, AtMan5-5 is highly expressed in mature dry seeds, whereas both genes are induced by germination [31]. Knockout analysis shows that AtMan5-5, but not AtMan5-2, is involved in seed germination [31]. Our extensive analysis of microarray expression databases indicates that AtMan5-2 and AtMan5-5 display opposing expression patterns during seed development and germination, where AtMan5-2 is highly expressed during stages 3 to 7 in the developing seed, and AtMan5-5 during stages 7–10 (Fig. 7a). Thus, the AtMan5-2 transcript is predominant in the immature seed, whereas the AtMan5-5 transcript is predominant in the mature dry seed. However, after 24 h imbibition, AtMan5-2 expression is re-induced, and it becomes the predominant transcript again. Furthermore, AtMan5-2 is highly expressed in the seed coat of heart stage embryo seeds, whereas AtMan5-5 is predominantly expressed in the contracted endosperm of mature embryo green seeds (Fig. 7b). Finally, AtMan5-2 is also highly expressed in the stem, and, similarly, the poplar single-gene AtMan5-2 ortholog, POPTR_0013S13400, is highly expressed in xylem (Fig. 7c), suggesting that AtMan5-2 subclade genes are members of the secondary cell wall gene program.

To further investigate whether AtMan5-2 subclade genes may be regulated by the secondary cell wall gene program, we examined if transcriptional cis-regulatory motifs in secondary cell wall synthesis are conserved in their upstream regions, since uncovering conserved cis-element motifs through bioinformatics allows elucidation of regulatory networks [59]. We focused on three cis-elements that mediate gene transactivation by transcription factors with well-established roles in secondary cell wall formation in the xylem: secondary wall NAC binding element (SNBE; [60]), which is bound by the NAC-type transcription factors NST1, SND1 and VND7; tracheyce element-regulating cis-element (TERE; [61]), which is a possible binding site of VND7; and the AC-element as
Fig. 6. (a) Ribbon representation of the AtMan5-2 homology model generated by SWISS-MODEL (α-helices are shown in pink, β-strands in blue and loops in white). (b) Semitransparent molecular surface that highlights the groove that runs across the protein molecule (same view as in panel a). (c) Close-up view of the mannoheptaose (yellow) modeled in AtMan5-2 across the groove. The docked mannoheptaose molecule is superimposed on the parental oligosaccharides from crystal-structure complexes: mannobiose (blue) in +1 to +2 from TrMan5A (PDB code 1QN8; [40]); mannotriose (red) in –1 to +2 from RmMan5B (PDB code 4LYQ; [41]); and mannotriose (green) in –4 to –2 from TmMan (PDB code 3MAN; [42]). (d) The mannoheptaose molecule (green) with individual subsites numbered from the aglycone end (–4) to the glycone end (+3). The non-reducing and reducing end of the oligosaccharide is denoted NR and R, respectively. Relevant protein side chains are shown. (e) Overlay of the structure template LeMan4a (blue; PDB code 1RH9) with the AtMan5-2 model (red), and the docked mannoheptaose molecule (yellow) across the groove. (f)–(i) Details of the individual subsites proposed in panel d: (f) subsites –4 and –3; (g) subsites –2 and –1; (h) subsites –1 and +1; (i) subsites +2 and +3.
levels
subclade
a
the
Fig.
the
recently
of
gest
cell
high
At
21
27)
AtMan5-5-5
involvement
TERE
transcription
expression
of
species
At
AtMan5-2
SNBE
motif
present
by
the
dicot
expression
is
for
the
conserved
seed
in
the
dicots.
AtMan5-5
also
expressed
in
a
conserved
expression
in
the
at
coincident
expression
in
the
xylem-
pal
endo
endosperm.
However,
AtMan5-2
subclade
genes
are
not
strictly
xylem-
specific,
as
they
are
also
expressed
in
seed
tissues.
In
the
developing
seed,
the
high
expression
of
AtMan5-2
in
the
heart-stage
embryo
seed
coat
is
coincident
with
mucilage
production
[63],
suggesting
a
role
for
AtMan5-2
in
seed
coat
mucilage
formation,
whereas
the
high
expression
of
AtMan5-5
in
the
contracted
endosperm
suggests
its
involvement
in
endosperm
hydrolysis.
Consistent
with
this,
knockout
of
AtMan5-5
delays
germination
time,
whereas
knockout
of
AtMan5-2
does
not
[31].
Glucomannan,
which
is
the
preferred
substrate
for
AtMan5-2,
has
been
identified
in
the
seed
mucilage
[12],
and
also
in
the
Arabidopsis
stem
[8].
Thus,
a
possible
biological
function
of
the
two
AtMan5-2-cluster
enzymes
may
be
to
catalyze
 glucomannan
hydrolysis
in
these
tissues,
but
whether
the
activity
is
required
for
structural
remodeling
of
the
cell
wall,
or
for
generating
manno-oligosaccharides
as
signal
molecules,
or
both,
cannot
be
concluded
from
the
present
data.
It
is
intriguing
that
genes
coding
for
AtMan5-2
cluster
enzymes
have
undergone
a
gene
duplication
event
early
in
Brassicaceae
evolution,
and
that
the
two
genes
display
opposing
gene
expression
patterns
in
seed
tissues.
Here,
AtMan5-2
and
AtMan5-5
could
possi-
bly
participate
in
bidirectional
cross-talk
between
seed
coat
and
endosperm,
a
process
where
these
two
tissues
mutually
affect
each
other’s
development
(reviewed
by
[64]).
However,
no
seed
coat
phenotype
was
reported
in
knockout
studies
of
AtMan5-2
and
AtMan5-5
[31],
so
such
possibility
should
necessarily
involve
redundant
functions
in
other
genes.

Fig. 7. Expression
of
AtMan5-2
and
AtMan5-5,
and
of
the
poplar
ortholog
POPTR_0013s13400.
Values
were
obtained
from
the
BAR
eFP
microarray
database.
(a) Expression
levels
of
AtMan5-2
(At2g20680)
and
AtMan5-5
(At4g28320)
during
the
Arabidopsis
life
cycle.
Seeds
stage
3–5
are
with
siliques.
(b) Expression
levels
of
AtMan5-2
(At2g20680)
and
AtMan5-5
(At4g28320)
in
the
seed.
(c) Expression
levels
of
POPTR_0013s13400
in
poplar
tissues.

recently
defined
by
Ratke
et
al.
[62],
which
is
the
binding
site
of
the
transcription
factor
MYB46.
Using
MEME,
a
particular
variant
of
the
SNBE
motif
that
is
strikingly
conserved
in
dicot
AtMan5-2
subclade
genes
(Table
2)
was
identified.
Although
several
(ten
out
of
27)
species
lack
this
SNBE
variant,
they
display
other
variants
of
the
SNBE
(not
shown),
and
thus
SNBE
motifs
are
highly
conserved
in
AtMan5-2
subclade
genes.
The
analysis
also
established
that
the
AC
motif
is
present
in
the
proximal
promoter
region
(700
bp
from
ATG)
in
21
out
of
27
dicot
species
(Supplementary
Table
S1).
In
contrast,
the
TERE
motif,
which
is
related
to
tracheary
element
programmed
cell
death
[61]
is
not
conserved
in
Man5-2
genes.
These
results
suggest
that
expression
in
the
xylem
is
a
conserved
feature
of
AtMan5-2
subclade
genes
in
dicots.

However,
AtMan5-2
subclade
genes
are
not
strictly
xylem-
specific,
as
they
are
also
expressed
in
seed
tissues.
In
the
developing
seed,
the
high
expression
of
AtMan5-2
in
the
heart-stage
embryo
seed
coat
is
coincident
with
mucilage
production
[63],
suggesting
a
role
for
AtMan5-2
in
seed
coat
mucilage
formation,
whereas
the
high
expression
of
AtMan5-5
in
the
contracted
endosperm
suggests
its
involvement
in
endosperm
hydrolysis.
Consistent
with
this,
Table 2
An SNBE-like motif is conserved in dicot Man5-2 genes.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Strand</th>
<th>Pos*</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesculenta_cassava1.02270753m</td>
<td>+</td>
<td>-571</td>
<td>AGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Mesculenta_cassava1.00340808m</td>
<td>-</td>
<td>-550</td>
<td>AGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Rccomminus_29983.0000804</td>
<td>+</td>
<td>-587</td>
<td>GAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Lusitatisitmus_Act10033687.g</td>
<td>+</td>
<td>-187</td>
<td>GCAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Lusitatisitmus_Act10031650.g</td>
<td>+</td>
<td>-186</td>
<td>GCAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Lusitatisitmus_Act10039833.c</td>
<td>+</td>
<td>-269</td>
<td>GCAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Lusitatisitmus_Act10018595.c</td>
<td>+</td>
<td>-241</td>
<td>GCAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Pirhocastra_Petri_01G130400</td>
<td>+</td>
<td>-588</td>
<td>GAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>P vulgaris_Pvulv.011G118900</td>
<td>+</td>
<td>-544</td>
<td>GAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Gmax_v1.1_Glyma1212770</td>
<td>+</td>
<td>-552</td>
<td>GAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Gmax_v1.1_Glyma0644750</td>
<td>+</td>
<td>-572</td>
<td>GAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Ppescia_pp0005783m.g</td>
<td>+</td>
<td>-544</td>
<td>GAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Mdmesquite_MDP0000709705</td>
<td>+</td>
<td>-525</td>
<td>GAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Mdmesquite_MDP0000417632</td>
<td>+</td>
<td>-546</td>
<td>GAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Fvesca_gene17714.v1.0-hybrid</td>
<td>+</td>
<td>-413</td>
<td>CCBATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Brapa_Chipu401_v1.2_Br011137</td>
<td>-</td>
<td>-628</td>
<td>CCBATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Grain mondio_Gorai.013075700</td>
<td>+</td>
<td>-433</td>
<td>CCBATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Grain mondio_Gorai.0050201100</td>
<td>+</td>
<td>-375</td>
<td>CCBATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Tcacao_Thece1EG008116</td>
<td>-</td>
<td>-361</td>
<td>CCACTGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Csinesis_orange1.1g14151m.g</td>
<td>+</td>
<td>-65</td>
<td>AAGACCTATGATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Csinesis_orange1.1g14151m.g</td>
<td>+</td>
<td>-65</td>
<td>AAGACCTATGATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Celemelina_Ciclev10015394am</td>
<td>+</td>
<td>-52</td>
<td>AAGACCTATGATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Celemelina_Ciclev10015394am</td>
<td>+</td>
<td>-52</td>
<td>AAGACCTATGATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Egrandis_Eucgr.A01823</td>
<td>-</td>
<td>-709</td>
<td>CCBATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Vvinifera_GSIVG01020553001</td>
<td>+</td>
<td>-628</td>
<td>GAGATGAATATTTTCCAGGATG</td>
</tr>
<tr>
<td>Mguittatua_v1.1_mgL007803m.g</td>
<td>-</td>
<td>-366</td>
<td>CCBATGAATATTTTCCAGGATG</td>
</tr>
</tbody>
</table>

CONSENSUS

SNBE

*Indicates distance in bp to translational start.

References


