A β-L-Arabinopyranosidase from Streptomyces avermitilis Is a Novel Member of Glycoside Hydrolase Family 27*

Hitomi Ichinose†1, Zui Fujimoto§1, Mariko Honda‡, Koichi Harazono¶, Yukifumi Nishimoto*, Atsuko Uzura*, and Satoshi Kaneko*12

From the *Food Biotechnology Division, National Food Research Institute, 2-1-12 Kannondai, Tsukuba, Ibaraki 305-8642, the †Protein Research Unit, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, and the §Research & Development Center, Nagase & Company, Limited, 2-2-3 Murotani, Nishi-Ku, Kobe, Hyogo 651-2241, Japan

Arabinogalactan proteins (AGPs) are a family of plant cell surface proteoglycans and are considered to be involved in plant growth and development. Because AGPs are very complex molecules, glycoside hydrolases capable of degrading AGPs are powerful tools for analyses of the AGPs. We previously reported such enzymes from Streptomyces avermitilis. Recently, a β-L-arabinopyranosidase was purified from the culture supernatant of the bacterium, and its corresponding gene was identified. The primary structure of the protein revealed that the catalytic module was highly similar to that of glycoside hydrolase family 27 (GH27) α-D-galactosidases. The recombinant protein was successfully expressed as a secreted 64-kDa protein using a Streptomyces expression system. The specific activity toward p-nitrophenyl-β-L-arabinopyranoside was 18 µmol of arabinose/min/mg, which was 67 times higher than that toward p-nitrophenyl-α-D-galactopyranoside. The enzyme could remove 0.1 and 45% L-arabinose from gum arabic or larch arabinogalactan, respectively. X-ray crystallographic analysis reveals that the protein had a GH27 catalytic domain, an antiparallel β-domain containing Greek key motifs, another antiparallel β-domain forming a jellyroll structure, and a carbohydrate-binding module family 13 domain. Comparison of the structure of this protein with that of α-D-galactosidase showed a single amino acid substitution (aspartic acid to glutamic acid) in the catalytic pocket of β-L-arabinopyranosidase, and a space for the hydroxymethyl group on the C-5 carbon of D-galactose bound to α-galactoside was changed in β-L-arabinopyranosidase. Mutagenesis study revealed that the residue is critical for modulating the enzyme activity. This is the first report in which β-L-arabinopyranosidase is classified as a new member of the GH27 family.

Arabinogalactan proteins (AGPs) are a family of complex proteoglycans widely distributed in plants (1, 2). AGPs are also found in tree exudate gums and coniferous woods (3) and are characterized by the presence of large amounts of carbohydrate components rich in galactose (all the sugars in the present study are in the D-configuration unless otherwise specified) and L-arabinose and by protein components rich in hydroxyproline, serine, threonine, alanine, and glycine (4). Type II arabinogalactans and short oligosaccharides are the two types of carbohydrates attached to the AGP backbone. Type II arabinogalactans have β-1,3-linked galactosyl backbones in mono- or oligo-β-1,6-galactosyl and/or L-arabinosyl side chains (2, 5). L-Arabinose and lesser amounts of other auxiliary sugars such as glucuronic acid, l-rhamnose, and l-fucose are attached to the side chains primarily at nonreducing termini (2). Molecular and biochemical evidence indicates that AGPs have specific functions during root formation, promotion of somatic embryogenesis, and attraction of pollen tubes to the style (6). However, because many putative protein cores exist and the structures of the carbohydrate moieties are complex, it has been difficult to differentiate one AGP species from another in plant tissues. This, in turn, has made it difficult to assign specific roles to individual AGPs. Despite significant physiological interest in AGPs, there are few studies on glycoside hydrolases that cleave the sugar moieties of these proteins. It is important to study such enzymes because hydrolytic enzymes specific to particular sugar residues or to a type of glycosidic linkage would be useful tools in the structural analysis of AGPs.

So far, we have focused on the β-1,3-β-1,6-galactan backbone, which is the common structure of heterogeneous AGPs, to identify glycoside hydrolases acting on AGPs. Galactanases that hydrolyze β-1,3- or β-1,6-galactosyl linkages are useful tools because the enzymes hydrolyze AGPs and produce the constituent carbohydrate moieties of AGPs. We cloned two kinds of galactanases: exo-β-1,3-galactanase (EC 3.2.1.145) from Phanerochaete chrysosporium and endo-β-1,6-galactanase (EC 3.2.1.164) from Trichoderma viride, and demonstrated that the enzymes were novel and could be classified as glycoside hydrolase family 43 (GH43) and family 5 (GH5), respectively (7–9) (see the CAZy website). Genes encoding proteins similar to such enzymes were also identified in the Streptomyces avermitilis genome (10, 11).

Because S. avermitilis has two different kinds of galactanases, we focused on finding novel AGP-degrading enzymes. We have cultivated the actinomycete using gum arabic as a carbon source, and isolated a novel β-L-arabinopyranosidase.
β-L-Arabinopyranosidase from *S. avermitilis*

To the best of our knowledge, the only report on β-L-arabinopyranosidase (EC 3.2.1.88) has been on its purification from *Cajanus indicus* (12). The amino acid composition of the enzyme was investigated (13), but its sequence remains unknown. In this article, we cloned β-L-arabinopyranosidase from *S. avermitilis* (SaArap27A), analyzed its catalytic properties, and analyzed the crystal structure of the recombinant enzyme. The results clearly showed that this enzyme is β-L-arabinopyranosidase and is a novel member of the glycoside hydrolase family 27 (GH27). This is the first detailed report on β-L-arabinopyranosidase.

**EXPERIMENTAL PROCEDURES**

**Substrates**—p-Nitrophenyl glycosides (PNP glycosides) (such as PNP-β-L-arabinopyranoside (PNP-β-L-Arap), PNP-α-L-arabinopyranoside, PNP-α-L-arabinofuranoside; PNP-α-L-galactopyranoside (PNP-α-Galp), PNP-β-L-galactopyranoside, PNP-α-glucopyranoside, PNP-β-glucopyranoside, PNP-α-mannopyranoside, PNP-β-mannopyranoside, PNP-α-xylopyranoside, PNP-β-xylopyranoside, PNP-α-L-fucopyranoside, PNP-β-L-fucopyranoside, PNP-α-L-rhamnopyranoside, PNP-β-L-rhamnopyranoside, PNP-β-galactouronide, PNP-β-glucuronide, PNP-α-N-acetylgalactosaminide, PNP-β-N-acetylgalactosaminide, PNP-α-N-acetylgalcosaminide, and PNP-β-N-acetylgalcosaminide), larch arabinogalactan, guar gum, and locust bean gum were purchased from Sigma. Debranched arabinan and wheat arabinoxylan were from Megazyme International (Wicklow, Ireland). Gum arabic was obtained from Nacalai Tesque, Inc. (Kyoto, Japan). Corn hull arabinoxylan was from Nihon Syokuhin Kakoh (Fuji, Japan). Arabinan was prepared as previously (7).

The sugar composition of the polysaccharides was analyzed as described previously (14, 15). Briefly, the substrate was hydrolyzed with 2 M trifluoroacetic acid by incubating at 121 °C for 1 h, and the acid was evaporated under a stream of N2 gas. The sugar composition of the sample was analyzed by high performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) using a CarboPac™ PA20 column ( Dionex Corp., Sunnyvale, CA) at a flow rate of 0.5 ml/min, as reported earlier (15).

**Partial Purification of Native SaArap27A**—*S. avermitilis* NBRC14893 was obtained from the National Institute of Technology and Evaluation (Kazusa, Japan). The strain was grown on liquid medium containing 1% gum arabic, 0.1% yeast extract, 0.25% glucose, 0.2% ammonium sulfate (11) in a baffled flask at 28 °C for 6 days. The culture was filtered, and the supernatant was used as the crude enzyme.

The crude enzyme was concentrated 10-fold using a 100-kDa polyethersulfone ultrafiltration membrane (Biomax, Millipore Corp.). Ammonium sulfate was added with stirring to the concentrated fraction to 70% saturation. The precipitated protein was redissolved in 50 mM acetate buffer (pH 4.0) containing 2 mM CaCl2 and dialyzed against the same buffer. After centrifugation, the supernatant was applied to a SP-Sepharose Fast Flow column (GE Healthcare UK Ltd; HR16/10 column, 0.8 × 10 cm) equilibrated with 50 mM acetate buffer (pH 4.0) containing 2 mM CaCl2. β-L-Arabinopyranosidase activity was eluted with a linear gradient of 0–1 M sodium chloride. The active fractions were combined, dialyzed against 50 mM acetate buffer (pH 4.0) containing 2 mM CaCl2, and loaded onto a Mono S HR 5/5 cation-exchange column (GE Healthcare) equilibrated with the same buffer. β-L-Arabinopyranosidase activity was eluted by a linear gradient of 0–0.5 M sodium chloride. The pooled active fractions were assayed for total protein and β-L-arabinopyranosidase activity, and protein purity was determined by SDS-PAGE, according to the method of Laemmli (16).

**Molecular Cloning, Mutagenesis, and Expression of Recombinant SaArap27A**—The pooled fractions from Mono S cation-exchange chromatography were separated by SDS-PAGE and electroblotted onto a polyvinylidene difluoride membrane (Millipore Corp.). The N-terminal amino acid sequence was analyzed on an HP G105A protein sequencer (Hewlett Packard).

The gene encoding a putative β-L-arabinopyranosidase (SAV_2186; GenBank™ accession number BAC69897) was cloned and expressed as a mature protein using the Streptomyces expression system as described previously (17). Briefly, the full-length gene amplified from *S. avermitilis* genomic DNA by PCR was cloned into a Streptomyces expression vector (18), and the plasmid was transformed into *S. lividans* 1326, which served as the host. The recombinant protein SaArap27A was secreted at high levels in the culture broth and was purified on lactosyl-Sepharose as described previously (19). The protein was detected by SDS-PAGE and dialyzed against 50 mM sodium phosphate buffer (pH 7.0). The final preparation thus obtained was used as the purified enzyme. The E99D mutant of SaArap27A was generated by PCR using appropriate primers (supplemental Table S1). Mutation was confirmed by DNA sequencing. Expression and purification of the mutant were carried out in the same way as for the wild-type enzyme.

**Enzymatic Properties**—β-L-Arabinopyranosidase activity was determined using a mixture containing 25 μl of 2 mM PNP-β-L-Arap, 20 μl of McIlvaine buffer (0.2 M NaHPO4 and 0.1 M citric acid, pH 4.0), and 5 μl of enzyme solution. The reactions were carried out at 40 °C for 10 min and terminated by the addition of 50 μl of 0.2 M Na2CO3. The amount of PNP released was detected at 400 nm (extinction coefficient = 19,200 M−1 cm−1). One unit of enzyme activity is defined as the amount of enzyme that released 1 μmol of PNP per minute. The protein amounts were determined by the BCA protein assay reagent kit (Pierce) using bovine serum albumin as the standard.

The effects of pH and temperature on enzyme activity were investigated as described previously (7, 20). The following buffers were used to study the effect of pH on β-L-arabinopyranosidase activity: McIlvaine buffer (pH 2.6–7.6), Atkins-Pantin buffer (pH 7.6–11.0), and glycine-HCl buffer (pH 1–2.6). The activity of β-L-arabinopyranosidase was assayed under the conditions described above. The effects of metal ions and chemicals on enzyme activity were also examined by treatment with 2 mM solutions of metal ions, including Ca2+, Cd2+, Co2+, Cu2+, Fe2+, Mg2+, Mn2+, Zn2+, Hg2+, and Ag+, and with 2 mM solutions of chemicals such as EDTA, p-chloromercuribenzoic acid, and SDS.

**Substrate Specificity**—The substrate specificity of SaArap27A toward various PNP glycosides was determined. The assay method was identical to that described for PNP-β-L-Arap. The kinetic parameters of wild-type and mutants of SaArap27A
were determined as follows. The reactions were performed in McIlvaine buffer (pH 4.0) containing 0.5–5 mM substrates, 0.1% (w/v) bovine serum albumin, and 2.6–260 nM enzyme at 37 °C for up to 15 min. The amount of PNP released was determined from the absorbance at 400 nm. The assay was performed in triplicate.

The substrate specificity of SaArap27A toward polysaccharides was also determined at 40 °C in McIlvaine buffer (pH 4.0) with 0.5% (w/v) polysaccharide as the substrate and 2.6 μM enzyme. After incubation for the appropriate reaction time, the initial hydrolysis rate was determined by the Somogyi-Nelson method (21). After incubation for 24 h, the amount of released arabinose was quantified by the HPAEC-PAD system using a CarboPac™ PA1 column (Dionex Corp.) at a flow rate of 1 ml/min, as described previously (20). The substrate hydrolysis rate was estimated by assuming that 100% l-arabinose was released from the substrate.

**RESULTS**

**Purification of Native SaArap27A and Analysis of the Primary Structure**—S. avermitilis was cultivated in liquid medium using gum arabic as the sole carbon source. The enzyme activities of the culture supernatant were tested using various PNP glycosides (see supplemental Fig. S1A). Interestingly, the culture supernatant showed the highest activity toward PNP-l-L-Arabinopyranosidase (658 amino acids) (Table 1). After cation-exchange chromatography, the enzyme appeared as a major band of a 64-kDa protein on SDS-PAGE (supplemental Fig. S2, lane 2, indicated by the arrow). The N-terminal amino acid sequence of the protein was determined to be AVTTR-QITVPSA. A BLASTP search against the S. avermitilis database revealed that the determined amino acid sequence corresponded to the open reading frame associated with the signal sequence predicted by SignalP.

**TABLE 1**

| Purification step                | Total volume | Total protein | Total activity | Specific activity | Purification | Yield |
|----------------------------------|--------------|---------------|---------------|------------------|--------------|-------|
| Crude enzyme                     | 4000         | 3440          | 169           | 0.049            | 1            | 100   |
| Ammonium sulfate precipitation   | 28           | 100           | 108           | 1.1              | 22           | 64    |
| SP-Sepharose Fast Flow          | 19           | 3.2           | 24            | 7.6              | 154          | 14    |
| Mono S                           | 5            | 1             | 23            | 21               | 429          | 12    |

The crystal structure was first analyzed by the molecular replacement method using the 1.5-Å resolution ligand-free data. Homology models of the first two domains and the C-terminal carbohydrate-binding module family 13 (CBM13) of SaArap27A were built with the MODELER program (Accelrys Software Inc., San Diego, CA) using the known crystal structures of rice α-galactosidase and CBM13 of Streptomyces olivaceoviridis β-xylanase (SoCBM13), respectively, as reference models (25, 26). The first run of the MOLREP program (27) in the CCP4 program suite (28) using the α-galactosidase-derived model as the reference against the native data resulted in two prominent solutions yielding an R factor of 0.553. This was followed by restrained refinement of the resultant models with the REFMAC5 program (29). With the refined models fixed, the second run of MOLREP using the CBM13-derived model as the reference yielded two solutions resulting in an R factor of 0.501. The second run of restrained refinement yielded an R factor of 0.433 and figure of merit of 0.521, and the resultant electron density map was sufficient for tracing the overall structure of two β-l-arabinopyranosidase molecules. The model was initially built using the automodeling program ARP/wARP (30), and several cycles of manual rebuilding and refinement followed using COOT (31) and REFMAC5 programs. In addition to the protein, models of water, glycerol, polyethylene glycol, sulfate ion, and HEPES were included. Structural analyses of the l-arabinose and galactose complexes were performed by isomorphous replacement, and the bound sugar molecules were identified by the F_{obs} – F_{calc} electron density map. The structure refinement statistics are shown in Table 3. The stereochemistry of the models was analyzed by the PROCHECK program (32). Figures were prepared using MOLSCRIPT (33) and RASTER3D (34).
structures have been determined, the deduced amino acid sequence corresponding to the catalytic module of SaArap27A (residues 45–345) resembled the sequences of the following proteins: rice α-galactosidase (41% identity and 56% similarity; Protein Data Bank entry 1UAS), chicken α-N-acetylgalactosaminidase (34% identity and 47% similarity; PDB entry 1R47), α-galactosidase from Hypocrea jecorina (32% identity and 45% similarity; PDB entry 15Z), and a putative α-N-acetylgalactosaminidase from Bacillus halodurans (22% identity and 36% similarity; PDB entry 3C1). In SaArap27A, two aspartic acids that are the catalytic residues of the GH27 enzymes are conserved (Fig. 1, indicated by the asterisk). Based on the crystal structures, almost all of the amino acids involved in sugar binding were conserved, but only one aspartic acid, which coordinates to the O6 atom of the bound sugar, was replaced by glutamic acid (Glu99) in SaArap27A.

On the other hand, CBM13 of SaArap27A (SaCBM13) was similar to certain structural elements in the following proteins whose three-dimensional structures have been elucidated: the xylan-binding domain chain A from S. olivaceoviridis (SoCBM13, 55% identity and 67% similarity; PDB entry 1XYF_A), xylan-binding domain chain A from S. lividans (SICBM13, 51% identity and 67% similarity; PDB entry 1KNL_A), ebulin chain B1 from Sambucus ebulus (34% identity and 54% similarity; PDB entry 1HWM_B), agglutinin chain B from Abrus precatorius (35% identity and 49% similarity; PDB entry 2Q3N_B), and ricin chain B from Ricinus communis (34% identity and 48% similarity; PDB entry 2AAL_B).

Expression and Characterization of SaArap27A—The DNA fragment encoding the full-length protein was cloned. Recombinant protein SaArap27A was successfully expressed in S. lividans in the secreted form and purified as a single band...
The enzyme was incubated at 30 °C for 1 h and was also stable between pH 4.0 and 8.0 at 30 °C for 1 h and was also stable at pH 4.0 and 40 °C. The enzyme was incubated at 40 °C for 24 h in a mixture containing 0.5% (w/v) substrate and McIlvaine buffer (pH 4.0).

Substrate specificity of SaArap27A toward polysaccharides
The enzyme activity (data not shown). The specific activity of SaArap27A toward polysaccharides was determined by the molecular replacement method at 1.5-, 1.9-, and 1.9-Å resolution, respectively, and the structures were refined to R/R_free factors of 0.157/0.179, 0.162/0.207, and 0.142/0.177, respectively (Table 3). The final models included two noncrystallographic symmetry-related SaArap27A molecules (molecules A and B) as well as the surrounding water, glycerol, polyethylene glycol, HEPES molecules, and sulfate ions. The SaArap27A molecule was composed of a single polypeptide chain of 614 amino acids (45–658), and molecule A was modeled through the chain, but the two N-terminal residues Ala6 and Val6 of molecule B could not be identified due to the lack of electron density. The root mean square difference of the Co atoms of these two molecules was calculated to be 0.31 Å, and two noncrystallographic symmetry molecules had almost the same overall structures.

The SaArap27A monomer consists of four domains (Fig. 2A). The N-terminal catalytic domain (domain I, residues 45–339) has a β/α barrel, which is observed in many glycoside hydrolases. The second domain (domain II, residues 340–430) is an eight-stranded anti-parallel β-domain containing tandemly repeated Greek key motifs, but in imperfect shapes. The relative arrangement of these two domains is the same as that of other GH27 enzymes. Domain II is located at the 7th and 8th α-helices of the catalytic domain. The third domain (domain III, residues 431–531) also contains eight antiparallel β-strands but comprises a β-jellyroll domain. This domain is located adjacent to domain II and also contacts the catalytic domain over the 5th α-helix and the loop after the 6th α-helix. The last domain, i.e., the C-terminal domain (domain IV, residues 532–658), is a ricin-type lectin domain consisting of the β-trefoil fold and is of the CBM13 type. This domain is in front of the catalytic domain covering the 6th α-helix and the loop before the 5th α-helix so that it contacts all three other domains forming a compact entity. No remarkable linker peptide could be observed between domains. The SaArap27A molecule contains four disulfide bonds: Cys148–Cys188 in the catalytic domain and Cys543–Cys562, Cys585–Cys604, and Cys628–Cys647 in CBM13.

Sugar complex crystals were prepared by soaking the SaArap27A crystals with l-arabinose or galactose solution, and the structures were determined. The relative positions of the domains did not change in comparison to the ligand-free state. The bound sugars were identified in the Fobs − Fcalc electron density maps (Fig. 3). In the l-arabinose complex, one l-arabinose molecule was bound in each catalytic domain and four (in molecule A) or three (in molecule B) l-arabinose molecules were found in the CBM domain (Fig. 2A). On the other hand, in

TABLE 2
Substrate specificity of SaArap27A toward polysaccharides

| Substrate | Sugar composition | Hydrolysis rate | mol % |
|-----------|------------------|-----------------|------|
| Gum arabic | Rha:Ara:Gal:GlcA = 40:56:11:10 | 0.1* | |
| Larch arabinoxylan | Ara:Gal = 10:51 | 45* | |

*The l-arabinose content in the residue after the enzymatic reaction was determined, and the hydrolysis rate was estimated by assuming that 100% arabinose was released from the substrate.

**FIGURE 1. Sequence alignment of SaArap27A with related proteins.** Domains I and II, comparison of the amino acid sequences of SaArap27A and GH27s. Shown is the alignment of deduced protein sequences of S. avermitilis SaArap27A (present study), O. sativa α-galactosidase (OsGal, PDB entry 1UAS), H. jecorina putative α-galactosidase (HjGal, 1SZN), Homo sapiens α-galactosidase (HsGal, PDB entry 1R47), and Gallus gallus α-N-acetylgalactosamidase (GgGalNA, PDB entry 1KTC). Secondary structural elements of SaArap27A are shown above the sequence. The N-terminal amino acid sequence determined is underlined. Catalytic residues are indicated by asterisks. The identified amino acid residues are enclosed in black boxes. The ligand-binding residues are enclosed in open boxes. Domain III, unknown domain of SaArap27A. Domain IV, sequence alignment of CBM13s, SaCBM13 (present study), SocCBM13 (PDB entry 1XYF_A), SicCBM13 (1KNL_A), ricin1, ricin chain B from R. communis (PDB entry 2AAI_B); ebublin1, ribosome-inactivating protein from S. ebulus (PDB entry 1HWM_B). Residues involved in substrate binding in SaCBM13 are indicated by open circles.

**β-l-Arabinopyranosidase from S. avermitilis**
β-1-Arabinopyranosidase from S. avermitilis

Substrate Binding Structure in the Catalytic Module—The structure of the λ-arabinose complex revealed that one λ-arabinose molecule was bound in the active site of the catalytic domain, and λ-arabinose was mainly in the α-anomeric form (Fig. 4A). Two aspartic acid residues, Asp186 and Asp247, are regarded as the catalytic residues of this enzyme. Asp186 forms a hydrogen bond from its Oδ2 atom to the λ-arabinose (Ara)-O1 atom, whereas Asp247 forms a hydrogen bond from its Oγ1 atom to the Ara-O1 atom. But if the bound λ-arabinose adopts the β-anomeric form as in the case of the natural substrate, the Asp247-Oγ1 atom is located at hydrogen bonding distance to the Ara-O1 atom, and these residues maintain the stereochemical arrangement that is often observed in the retaining glycosidases (Fig. 4A). Apart from the catalytic residues, many other amino acids participate in sugar binding. The Arg242-N71 and Cys224-S75 atoms hydrogen bond to the Ara-O2 atom. The Lys184-N4 atom forms two hydrogen bonds to the Ara-O3 and O4 atoms, and the Asp98-O4 atom forms hydrogen bonds to the Ara-O3 atom. The indole group of Trp63 has stacking contact with the flat face of λ-arabinose O3-C3-C4-C5. The disulfide bridge Cys148-Cys188, Trp226, and Met283 form part of the active site pocket. λ-Arabinose is bound in a manner quite similar to that in which galactose is bound by rice α-galactosidase (Fig. 4C). These residues are completely conserved in their positions and functions. At the entrance of the catalytic pocket, one glycerol molecule was observed (Fig. 4A). It was surrounded by Tyr152, Trp226, and Tyr250. The position seems to be the aglycon subsite of the enzyme. However, Tyr152 and Tyr250 are not conserved in the other GH27 enzymes because they are located in the inserted peptides unique to SaArap27A.

The structure of the galactose complex revealed that one galactose molecule was bound in the active site, and galactose was observed to mainly adopt the β-anomeric form (Fig. 4B). The B factors of the bound galactose atoms are comparable with those of the surrounding amino acids, and sugar occupancy is close to 100%. The position of the bound galactose is almost the same as that of bound λ-arabinose in the λ-arabinose complex (Fig. 4A), and the sugar-binding mechanism is also well conserved. However, a difference was observed in the case of the galactose (Gal)-O6 atom, which is not present in the λ-arabinose molecule. The Gal-O6 atom has a unique hydrogen bond to the Glu99-O4 atom. In comparison to the structure of rice α-galactosidase, there were no obvious differences in the galactose-binding manner, but Glu99, the hydrogen-bonding partner of the Gal-O6 atom, was replaced with Asp52 (Fig. 4C).

Role of Glu99 in SaArap27A to Modulate β-1-Arabinopyranosidase/α-D-Galactopyranosidase Activity—The Km value of SaArap27A for PNP-α-Galp was almost the same as for PNP-β-1-Arap (Table 4). However, the kcat values of the enzyme for PNP-α-Galp was ~140 times lower than for PNP-β-1-Arap. To investigate the role of Glu99 for enzyme activity, a mutant enzyme (SaArap27A/E99D) was constructed in which Glu99 was replaced by Asp. When the enzyme activities for PNP glycosides were tested (supplemental Table S2), the mutant gained higher activity for PNP-α-Galp than for PNP-β-1-Arap. The specific activity of SaArap27A/E99D for PNP-α-Galp was 9 units/mg at pH 4.0 at 40 °C, and was 9 times higher than that for PNP-β-1-Arap. The Km and kcat values of the mutant for PNP-α-Galp and PNP-β-1-Arap were 4.3 ± 0.1 mm and 29 ± 0.6

TABLE 3
Data collection and structure refinement statistics of SaArap27A
The values in parentheses represent the highest resolution shell.

| Data                  | Ligand-free | λ-Arabinose complex | D-Galactose complex |
|----------------------|-------------|---------------------|---------------------|
| Data collection      | P21_2_1     | P21_2_1             | P21_2_1             |
| Unit cell parameters (Å) | a = 68.2   | a = 68.3            | a = 68.3            |
|                      | b = 98.9    | b = 99.0            | b = 99.1            |
|                      | c = 181.3   | c = 181.3           | c = 181.6           |
| Beam line            | PF BL17     | PF BL6A             | PF BL6A             |
| Wavelength (Å)       | 0.98300     | 0.97800             | 0.97800             |
| Resolution (Å)       | 50.0-1.5 (1.55-1.50) | 50.0-1.9 (1.97-1.90) | 50.0-1.9 (1.97-1.90) |
| Rmerge (%)           | 0.086 (0.335) | 0.075 (0.230)       | 0.088 (0.286)       |
| Completeness (%)     | 98.1 (91.2) | 99.9 (100.0)        | 100.0 (100.0)       |
| Multiplicity         | 7.8 (8.3)   | 8.6 (8.7)           | 8.6 (8.7)           |
| Average I/σ(I)       | 30.6 (6.5)  | 31.7 (12.2)         | 28.4 (9.1)          |
| Unique reflections   | 191,711 (17,626) | 97,365 (9,588)    | 97,694 (9,630)      |
| Observed reflections | 1,599,786   | 874,401             | 835,515             |

Structure refinement

| Resolution (Å)       | 50.0-1.5 (1.55-1.51) | 50.0-1.9 (1.95-1.90) | 50.0-1.9 (1.95-1.90) |
| R factor             | 0.157 (0.217)        | 0.162 (0.226)        | 0.142 (0.152)        |
| Rfree factor         | 0.179 (0.223)        | 0.207 (0.295)        | 0.177 (0.182)        |
| RMSD from ideal value| Bond lengths (Å)     | 0.006                | 0.011                | 0.009                |
|                      | Bond angles (°)      | 1.066                | 1.212                | 1.120                |
|                      | No. of water molecules | 1827               | 1360                 | 1487                 |
|                      | Average B-value (Å²) | 13.5                 | 13.1                 | 13.4                 |

Ramachandran plot

| Most favored (%)     | 88.4                  | 89.0                  | 87.8                  |
| Additional allowed (%)| 11.0                  | 10.6                  | 11.6                  |
| Generously allowed (%)| 0.4                   | 0.3                   | 0.4                   |
| Disallowed (%)       | 0.2                   | 0.2                   | 0.2                   |

* Rmerge = ΣI(hkl) - Σ|I(hkl)| / ΣI(hkl), where I(hkl) is the ith observation of reflection hkl, and Σ|I(hkl)| is the weighted average intensity for all observations i of reflection hkl.
were bound in three subdomains of SaArap27A domain IV (Fig. 4, D–F). In subdomain α, the l-arabinose was docked with its hydrophobic face (C3·C4·C5) into the aromatic indole ring of Trp560 to form a stacking interaction (Fig. 4D). There are six hydrogen bonds between subdomain α and the bound l-arabinose molecule. These are from Asn548–O1 to the Ara-O1 atom; from Asn563–N81, Asn567–N83, and Asp545–O42 to the Ara-O3 atom; and from Asn548–N and Asp545–O41 to the Ara-O4 atom. Another hydrogen bond was observed from the side chain of Thr481 of domain III to the Ara-O1 atom. On the side of the binding pocket, the aromatic plane of the side chain of Tyr547 faces the C3·O3 bond of the bound l-arabinose. Behind Tyr547 of molecule A, another l-arabinose molecule was observed that forms a stacking interaction between the aromatic ring of Tyr547 and the hydrophobic surface of l-arabinose. Two carboxylate oxygen atoms of Glu558 form two hydrogen bonds with the O3 and O4 atoms of the l-arabinose. However, no bound l-arabinose was observed at this site in molecule B. The average B factor of this l-arabinose molecule was high in comparison to that of the bound l-arabinose in the binding pocket, and this binding did not seem to be specific.

In subdomains β and γ, the bound l-arabinose was docked in a manner similar to that in subdomain α (Fig. 4, E and F). The detailed binding manners of each subdomain are described under supplemental materials.

When the crystal was soaked with galactose, electron density in the sugar-binding sites of CBM13 showed that the bound molecules were mostly glycerol, which were derived from the crystallization condition. Additionally, the structure of the complex revealed that the CBM13 of SaArap27A bound to l-arabinose in three subdomains when the crystals were soaked in a mixture of l-arabinose and galactose.

**DISCUSSION**

β-l-Arabinopyranosidase was found in AGP as 3-O-β-l-arabinopyranosyl-l-arabinose determined by structural analysis of carbohydrate moieties (3, 35), implying the existence of a β-l-arabinopyranosidase. However, to the best of our knowledge, the only report on β-l-arabinopyranosidase has been on its purification from *C. indicus*, and its sequence information has not yet been elucidated (12, 13).
In this study, for the first time, we cloned the β-1-arabinopyranosidase gene from *S. avermitilis* and determined the three-dimensional structure of the enzyme. Sequence analysis showed that SaArap27A has a unique modular structure composed of four structural modules. The resultant three-dimensional structure showed that this enzyme has a compactly packed modular architecture in which all four domains are in contact with each other, and the catalytic pocket and three sugar-binding sites are open to the solvent. Interestingly, they face the same solvent region, *i.e.* in Fig. 2B, the catalytic site faces the left, and three l-arabinose-binding sites of CBM13 face upwards. We noticed that there are many tyrosine residues exposed to this solvent region. Tyr150, Tyr151, Tyr152, Tyr228, Tyr250, and Tyr251 are in the catalytic domain, and five of these are in the inserted peptides around the catalytic cleft (supplemental Fig. S5B). Tyr547, Tyr589, and Tyr632 are exposed from the CBM13 domain, and these residues were also used in the arabinose binding pocket mentioned under “Results” (Fig. 4, D–F). This solvent region seems a good environment for carbohydrate binding. The inserted peptides in the catalytic domain should relate to the binding of type II arabinogalactan. We mentioned that the function of domain III is unknown. However, it seems to play an important role as spacer domains together with domain II to allow closer positioning of the CBM13 to the catalytic site, although the domains themselves do not have sugar-binding activity.

The substrate type II arabinogalactan contains both l-arabinopyranosyl and l-arabinofuranosyl residues in addition to galactose. CBM13 is often represented by the plant lectin ricin; therefore, its main function is considered to be galactose binding. However, the structure of the sugar complex revealed that CBM13 of SaArap27A preferred to bind to l-arabinose when the crystals were soaked in a mixture of l-arabinose and galactose. The sugar binding pocket of the l, β, and γ subdomains of SaArap27A CBM13 bound l-arabinose in the same manner (Fig. 4, D–F). The affinity of CBM13 for l-arabinose was reported for SlCBM13 (36). However, the association constants of l-arabinose and galactose for SlCBM13 were almost the same, *i.e.* $1.5 \times 10^2$ and $6.1 \times 10^2$ M$^{-1}$, respectively (36). Although the structures of many CBM13 complexes are cur-

![FIGURE 4. A, structure of the catalytic pocket of SaArap27A in the l-arabinose complex structure. Hydrogen bonds between SaArap27A and bound l-arabinose (Arap) and glycerol (Gal) are shown as a dashed cyan line. B, structure of the catalytic pocket of SaArap27A in the galactose complex. C, superimposition of the catalytic pocket between SaArap27A (blue) and rice α-galactosidase (brown, PDB entry 1UAS). D, structure of the l-arabinose-binding pocket in the subdomain α of SaArap27A domain IV in the l-arabinose complex structure. Hydrogen bonds between SaArap27A and the bound l-arabinose are shown as a dashed cyan line. E and F, l-arabinose binding pocket in subdomains β and γ of SaArap27A domain IV in the l-arabinose complex structure.]

### TABLE 4

Activity of wild-type and mutant of SaArap27A toward PNP-β-1-Arap and PNP-α-Galp

| Enzyme             | PNP-β-1-Arap | PNP-α-Galp |
|--------------------|--------------|------------|
|                    | $k_{cat}$    | $K_m$      | $k_{cat}/K_m$ | $k_{cat}$    | $K_m$      | $k_{cat}/K_m$ |
| SaArap27A          | 317 ± 10     | 3.6 ± 0.4  | 88           | 2.3 ± 0.1    | 5.1 ± 0.3  | 0.5           |
| SaArap27A/E99D     | 17 ± 1       | 11.1 ± 0.9 | 1.5          | 29 ± 0.6     | 4.3 ± 0.1  | 6.7           |

The enzyme was incubated at 37 °C up to 15 min in a mixture containing 0.5–5 mM substrate and McIlvaine buffer (pH 4.0).
similar to SaArap27A, at the Gal-O6 position in the substrate with confirmed activity have aspartic acid, not glutamic acid, in Aspergillus nidulans. The sis study clearly demonstrated the critical role of Glu99 for the enzyme for PNP-β-L-AraGalp. The K_m value of SaArap27A for PNP-α-Galp was almost the same as that of PNP-β-L-Arap although the k_cat values of the enzyme were different (Table 4). The k_cat value of the enzyme for PNP-α-Galp was extremely lower than that for PNP-β-L-Arap, indicating that the galactose tightly bound to the catalytic pocket resulted in the reduction of turnover of the catalysis. Therefore, this substitution would be one factor why SaArap27A selects β-L-arabinopyranoside, not α-galactopyranoside, as its substrate. The results of the mutagenesis study clearly demonstrated the critical role of Glu99 for modulating the enzyme activity to possess β-L-arabinopyranosidase activity. The E99D mutant showed higher activity for PNP-α-Galp than for PNP-β-L-Arap (see Table 4 and supplemental Table S2).

The single aspartic acid to glutamic acid substitution is found in only SaArap27A, and Asp residues are conserved in all other GH27s whose three-dimensional structures are known, including α-galactosidases and α-N-acetylgalactosaminidases (Fig. 1). BLAST searches of the NCBI nonredundant protein sequence data base using the deduced amino acid sequence of SaArap27A revealed that the same substitution was also present in a putative α-galactosidase from S. sviceus ATCC 29083 (accession number EDY60900). The amino acid sequence shows 82% identity and 89% similarity. Additionally, the putative protein has a C-terminal module classified as a CBM13 domain. The putative protein also appears to be a β-L-arabinopyranosidase. The results of BLAST searches showed a different kind of amino acid substitution (aspartic acid-aspartic acid (Asp52 in rice α-galactosidase)-cysteine-tryptophan to aspartic acid-cysteine-glycine-tryptophan) in a putative protein annotated as an α-galactosidase. The putative α-galactosidase from Aspergillus nidulans (accession number ABF50881) was expressed in Pichia pastoris X-33 as a plant cell wall polysaccharide-degrading enzyme (37). The crude enzyme showed weak activity toward PNP-α-Galp (see Table 1 in Ref. 37). However, the enzyme did not show any activity toward raffinose, locust bean gum, and guar gum. A possible interpretation of these properties of the enzyme is that the real activity differs from that of α-galactosidase because all α-galactosidases with confirmed activity have aspartic acid, not glutamic acid, similar to SaArap27A, at the Gal-O6 position in the substrate binding pocket. Aspartic acid coordinating the Gal-O6 atom seems important for α-galactosidases or α-N-acetylgalactosaminidase activity and substitution of this aspartic acid would result in different substrate specificity. Similarly, we reported that some amino acid substitutions around the Gal-O2 atom differentiate the enzyme activity between α-galactosidase or α-N-acetylgalactosaminidases. These sugar complex structures of GH27s showed the possibility of alternation of substrate specificity by some amino acid mutations and gave a structural base for further molecular design of GH27 enzymes.

In conclusion, we purified an enzyme with β-L-arabinopyranosidase activity from S. avermitilis. The gene encoding the enzyme was cloned, and the amino acid sequence was determined. In addition, we successfully determined the intact three-dimensional structure. The protein is classified as GH27 and consists of four modules. The enzymatic activity differs from that of other α-galactosidases due to a single amino acid substitution. The structure of the L-arabinose complex clearly indicates that the enzyme is β-L-arabinopyranosidase. Furthermore, CBM13 of this enzyme has a novel L-arabinose binding property. This is the first report of β-L-arabinopyranosidase as a new member of the GH27 family and CBM13 as an L-arabinose-binding module.

Acknowledgments—Synchrotron radiation used for this work was approved by the Photon Factory Advisory Committee (proposal number 2007G6659). We thank the scientists and staff at the Photon Factory for their help.

REFERENCES

1. Fincher, G. B., Sawyer, W. H., and Stone, B. A. (1974) Biochem. J. 139, 535–545
2. Gaspar, Y., Johnson, K. L., McKenna, J. A., Bacic, A., and Schultz, C. J. (2001) Plant Mol. Biol. 47, 161–176
3. Aspinall, G. O. (1969) Adv. Carbohydr. Chem. Biochem. 24, 333–379
4. Knox, J. P. (1995) FEBS Lett. 371, 1004–1012
5. Goodrum, L. J., Patel, A., Leykam, J. F., and Kieliszewski, M. J. (2000) Biochemistry 54, 99–106
6. Seifert, G. J., and Roberts, K. (2007) Annu. Rev. Plant Biol. 58, 137–161
7. Ichinose, H., Yoshida, M., Kotake, T., Kuno, A., Ishikawa, K., Tsumuraya, Y., Samejima, M., Hirabayashi, I., Kobayashi, H., and Kaneko, S. (2005) J. Biol. Chem. 280, 25820–25829
8. Kotake, T., Kaneko, S., Kubomoto, A., Haque, M. A., Kobayashi, H., and Tsumuraya, Y. (2004) Biochem. J. 377, 749–755
9. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) Nucleic Acids Res. 37, D233–D238
10. Ichinose, H., Kotake, T., Tsumuraya, Y., and Kaneko S. (2006) Biosci. Biotechnol. Biochem. 70, 2745–2750
11. Ichinose, H., Kotake, T., Tsumuraya, Y., and Kaneko, S. (2008) Appl. Environ. Microbiol. 74, 2379–2383
12. Dey, P. M. (1973) Biochim. Biophys. Acta 280, 25820–25829
13. Dey, P. M. (1983) Biochim. Biophys. Acta 746, 8–13
14. Yang, P., Ichinose, H., Nakajima, M., Kobayashi, H., and Kaneko, S. (2006) Food Sci. Technol. Res. 12, 43–49
15. Yoshida, M., Liu, Y., Uchida, S., Kawarada, K., Utagami, Y., Ichinose, H., Kaneko, S., and Fukuda, K. (2008) Biosci. Biotechnol. Biochem. 72, 805–810
16. Laemmli, U. K. (1970) Nature 227, 680–685
17. Fujimoto, Z., Ichinose, H., Harazono, K., Honda, M., Uzura, A., and Kaneko, S. (2009) Acta Crystallogr. F Struct. Biol. Cryst. Commun. 65, 632–634
18. Hatanaka, T., Onaka, H., Arima, J., Uraji, M., Uesugi, Y., Usuki, H., Nishi-
29. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. D Biol. Crystallogr. 53, 240–255
30. Cohen, S. X., Morris, R. J., Fernandez, F. I., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kléwyg, G. J., and Perrakis, A. (2004) Acta Crystallogr. D Biol. Crystallogr. 60, 2222–2229
31. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132
32. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
33. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
34. Merritt, E. A., and Bacon, D. J. (1997) Methods Enzymol. 277, 505–524
35. Odonmazig, P., Ebringerová, A., Machová, E., and Alfréd, J. (1994) Carbohydr. Res. 252, 317–324
36. Boraston, A. B., Tomme, P., Amandaron, E. A., and Kilburn, D. G. (2000) Biochem. J. 350, 933–941
37. Bauer, S., Vasu, P., Persson, S., Mort, A. J., and Somerville, C. R. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 11417–11422