Chapter 5

Crystal structure of Endo-Xylogalacturonan Hydrolase from Aspergillus tubingensis

Henriëtte J. Rozeboom¹, Gerrit Beldman², Henk A. Schols² and Bauke W. Dijkstra²

¹ Laboratory of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands. ² Laboratory of Food Chemistry, Wageningen University, Bornse Weilanden 9, 6708 WG Wageningen, The Netherlands

FEBS Journal (2013) 280 (23), 6061-6069

Endo-xylogalacturonan hydrolase features a wide-open active site groove to accommodate its xylogalacturonan substrate. No specific xylose binding pockets are present that could hinder sliding movements of the substrate of this processive enzyme. A model of xylosylated tri-galacturonate in the active site explains the enzyme's preference for hydrolysis of the glycosidic bond linking two β-xylose substituted galacturonic acid residues.
ABSTRACT

Endo-Xylogalacturonan hydrolase is a member of glycoside hydrolase family 28 (GH28) that hydrolyzes the glycosidic bond between two β-xylose substituted galacturonic acid residues in pectin. Presented here is the X-ray crystal structure of the endo-xylogalacturonan hydrolase from Aspergillus tubingensis (XghA) at 1.75 Å resolution. The high degree of structural conservation in the active site and catalytic apparatus, compared with polygalacturonases, indicates that cleavage of the substrate proceeds in essentially the same way as found for the other GH28 enzymes. Molecular modeling of a xylosylated tri-galacturonate in the active site identified the amino acid residues involved in substrate binding. They border a substrate-binding cleft, that is much wider than in other polygalacturonases, and can accommodate xylosylated substrates. The most extensive interactions appear to occur at subsite +2, in agreement with the enzyme kinetics results, which showed enhanced activity on substrates with a xylose attached to the galacturonic acid bound at subsite +2.

INTRODUCTION

Pectin is a mixture of highly heterogeneous and branched polysaccharides that constitutes a major component of the primary cell wall of plants (Wong, 2008). The predominant use of pectins are in the food and cosmetic industries as gelling and thickening agents, and as stabilizers. Pectins also have positive effects on human health and have multiple biomedical uses (Jayani et al., 2005, Mohnen, 2008). The structure of pectin consists of ‘smooth’ regions of homogalacturonan, and ‘hairy or ramified’ regions of rhamnogalacturonan I, rhamnogalacturonan II, xylogalacturonan (XGA), apiogalacturonan, and polysaccharides comprising mostly neutral sugars (Wong, 2008). The most abundant pectin, homogalacturonan, is a linear homopolymer of α-1,4-linked D-galacturonic acid (GalA) residues, with the carboxylic acid groups methyl-esterified to various degrees up to 70–80%. Xylogalacturonan (XGA) is a polymer of GalA units in which the O-3 atoms of 20-100% of the GalA residues carry a β-(1,3)-linked xylose side chain (Wong, 2008, Coenen et al., 2007, Yapo, 2011). XGA from pea hull (Pisum sativum) contains mainly single terminal xylose residues and a few short 1,2-linked, or sometimes 1,3-linked xylosyl oligosaccharides attached to O-3 (Le Goff et al., 2001). In XGAs from pectin of apple (Malus domestica) and potato (Solanum tuberosum), 1,4-linked xylose disaccharides are observed (Zandleven et al., 2006), while in pectin from soybean (Glycine max) branched XGAs occur with both 1,4 and 1,2 linkages. (Nakamura et al., 2002)
Polygalacturonases (PGs) produced by plants are involved in the ripening process to soften and sweeten fruits. In contrast, bacterial and fungal polygalacturonases assist in the degradation process to liberate nutrients on which they can grow. The filamentous fungi of the genus *Aspergillus* produce many extracellular hydrolytic enzymes to degrade the pectin backbone (de Vries et al., 2001), and these have been classified in the pectin degrading glycoside hydrolase family 28 (GH28). GH28 includes endo-polygalacturonases, exo-polygalacturonases and rhamnogalacturonan hydrolase (RGase A) (Cantarel et al., 2009). All family GH28 members act with an inverting mechanism (Zandleven et al., 2005).

Among the GH28 enzymes, endo-xylogalacturonan hydrolase (XghA, E.C. 3.2.1.-) plays a unique role in the degradation of XGA. XghA is a processive enzyme that cleaves the β-xylose substituted GalA backbone of pectin in an endo-manner, with xylose substitution being a prerequisite (van der Vlugt-Bergmans et al., 2000), although activity has also been observed towards di-galacturonic acid of which only one residue is xylosylated (Zandleven et al., 2006, Zandleven et al., 2005, Beldman et al., 2003). Mature XghA is a 388-residue protein with ~22-27% sequence identity to polygalacturonases from other *Aspergillus* species, but the structural determinants of its typical xylogalacturonase activity have remained obscure. The four active site sequence motifs that are absolutely conserved in GH28 are also present in XghA (Armand et al., 2000, Markovič et al., 2001). The optimal activity of XghA was found between pH 3.0-3.5, and its optimal temperature was ~50 °C (Beldman et al., 2003). The enzyme appeared to be stable over a broad pH range (2.5-6.5).

Here we present the 1.75 Å crystal structure of endo-xylogalacturonan hydrolase, and compare it to other GH28 family members, revealing that it has a more open active site to accommodate the xylose residues. Modeling of a bound (XGA)₃ molecule suggests that the most extensive interactions with xylose occur at subsite +2, in agreement with previous kinetic studies.

### RESULTS AND DISCUSSION

**Overall structure of endo-xylogalacturonan hydrolase**

The crystal structure of endo-xylogalacturonan hydrolase (XghA) was determined to a resolution of 1.75 Å using molecular replacement. Data collection and phasing are summarized in Table 1. The asymmetric unit contains one XghA monomer, one glycerol molecule, seven sulfate ions, and 458 water molecules, and yielded a final $R/ R_{free}$ of 16.7/18.9 % after refinement. Further details of the refinement are shown in Table 1. The N-terminal signal peptide (18 residues) and the first 16 residues of the mature
enzyme were not visible in the electron density maps. Thus, the first residue observed in the electron density map is Ala35.

**Table 1.** Data collection and refinement statistics. Values in parentheses are for the highest resolution shell.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffraction data</td>
<td>Native</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9724</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>40.7–1.75</td>
</tr>
<tr>
<td>Cell dimensions (Å) a, b, c</td>
<td>76.4, 121.2, 129.7</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>60896 (8795)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (100.0)</td>
</tr>
<tr>
<td>Overall I/σ (I)</td>
<td>11.6 (3.7)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt; (%)</td>
<td>10.4 (54.2)</td>
</tr>
<tr>
<td>R&lt;sub&gt;pim&lt;/sub&gt; (%)</td>
<td>4.2 (21.4)</td>
</tr>
<tr>
<td>R/ R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>16.7/18.9</td>
</tr>
</tbody>
</table>

R.m.s. deviations from ideal values

| Bond lengths (Å)          | 0.008            |
| Bond angles (°)           | 1.327            |
| Protein residues          | 35 - 406         |
| Water molecules           | 458              |
| Sulfate molecules         | 7                |
| Glycerol molecule         | 1                |
| PDB accession ID          | 4C2L             |

XghA folds into a right-handed parallel β-helical structure comprising ten complete turns (Fig. 1), which contain three twisted parallel β-sheets (PB1, PB2a, PB2b and PB3; names are as adopted for the PG2 polygalacturonate hydrolase (van Santen et al., 1999). The β-strands in the parallel β-sheets are connected by four turns, T<sub>1</sub>, T<sub>2a</sub>, T<sub>2b</sub> and T<sub>3</sub>. All T<sub>2a</sub> and T<sub>2b</sub> turns are short, whereas the lengths of the T<sub>1</sub> turns in the C-terminal part (from residue 259) and those of the T<sub>3</sub> turns in the N-terminal part (until residue 207) are extended. Three short α-helices are observed in the turns, α2 (residues 102-105) in the T<sub>3</sub> turn of coil 1, α3 (residues 136-144) in the T<sub>3</sub> turn of coil 2, and α4 (residues 326-331) in the T<sub>1</sub> turn of coil 9. Three incomplete coils are found at the N- and C-terminal ends (indicated as a–c in Fig. 1B).

The C-terminus of XghA (residues 392–406) makes a very tight turn and folds back on the protein, forming interactions with PB1 of coils 9, 10 and c (Fig. 2B). This turn is
stabilized by a disulfide bridge between Cys389 and Cys400, and by the C-terminal glycine residue, whose carboxylate oxygen atoms form hydrogen bonds with the main-chain nitrogen atoms of Ser320 and Cys321, as well as by many other hydrogen bonds. This seals off the hydrophobic interior of the β-helix at the C-terminus, while the N-terminus is capped by β1, directly followed by α1 (residues 50-60), similar to what has been observed for endo-PG1 (van Pouderoyen et al., 2003).

**Figure 1.** Crystal structure of the xylogalacturonan hydrolase XghA (N-terminus on the top, C-terminus on the bottom).

A. XghA is displayed in cartoon format and colored in a rainbow gradient from blue, to red. The O-linked glycosylation site is found at T38, whereas the N-glycosylation sites are found at N278 and N301.

B. Overview of the parallel β-sheets PB1, PB2a, PB2b and PB3, shown in yellow, orange, red and blue, respectively. The six disulfide bridges are shown in green.

C. APBS (Adaptive Poisson-Boltzmann Solver) electrostatic surface potential of XghA at pH4, highlighting the prominently positively charged active site cleft. The residues colored in red are negative, the ones colored in blue are positive and the ones in white are neutral. The negatively charged residues in the active site are masked by the surplus of positive amino acids and the low pH. The electrostatic surface has been homogeneously scaled between ±20 kT/e.
Six disulfide bridges are present: Cys37-Cys60 (1), Cys85-Cys88 (2), Cys230-Cys247 (3), Cys321-Cys329 (4), Cys363-Cys369 (5) and Cys389-Cys400 (6) (Fig. 1B and Fig.S1). Four of them (bridges 1, 3, 5 and 6) are conserved among fungal endo-PGs and RGase A (Fig. S1). Cys230, of the third disulfide bridge, follows directly the two catalytic aspartates Asp228 and Asp229 in the amino acid sequence, possibly to ensure correct local folding (Petersen et al., 1997). The fourth disulfide bridge stabilizes the T₁ turn of coil 9; one partner, Cys321, precedes the substrate-binding residue Tyr322 (see below), and may be important for stabilizing the position of this tyrosine. In addition, the backbone amide of Cys321 interacts with the carboxylate oxygen atoms of the C-terminal glycine residue (Fig. 2B).

A comparison of XghA with true polygalacturonases shows that XghA has the smallest root mean square difference (on Cα positions) with *Aspergillus niger* PG1 (van Pouderoyen et al., 2003), *Colletotrichum lupini* PG1 (Bonivento et al., 2008), *Aspergillus niger* PG2 (van Santen et al., 1999) and *Erwinia carotovora* polygalacturonase (Pickersgill et al., 1998), all within 1.7 Å.

**Glycosylation**

The molecular masses observed in the gel filtration chromatography and SDS-PAGE analyses are significantly higher than the 45 kDa predicted from the amino acid sequence. Glycosylation may cause a higher apparent molecular mass in gel filtration experiments, as well as in SDS-PAGE by decreased binding of SDS to the carbohydrate, compared to polypeptides of the same mass (Leach et al., 1980).

An N-linked N-acetylglucosamine (GlcNAc) residue was modeled in electron density extending from the Asn278 side chain. Similarly, at Asn301, two GlcNAc residues were visible. Additional sugar residues are probably present at these sites, but their electron densities were not sufficiently clear to model them. Furthermore, an O-glycosylation site

**Figure 2.** Close-up view of (A) the N- and (B) the C-terminus of XghA.
was found at Thr38 (Fig. 2A). At this position, the electron density matches that of a covalently bound mannose residue. In fungi, O-glycosylation has a role in modulating the function of secretory proteins by enhancing their stability and solubility. The O-mannosylation at Thr38 probably affords protection from degradation by proteases (Goto, 2007). The disulfide bridge between Cys37 and Cys60 may have a similar function.

**Active site**

The active site is located in an open groove/cleft on the surface of the protein. At the entrance to the active site, the width of the groove between T3 Ser146 and T1 Gly323 is significantly larger (~ 15.3 Å) than in other GH28 endo-PGs (in which it ranges from 6.4 Å in endo-PG1 to 11.0 Å in endo-PG2 (Fig. 1) (van Pouderoyen et al., 2003, Bonivento et al., 2008), suggesting easy access for xylosyl side chains along the length of pectin polymer (Fig. 1B). The bottom of the groove contains mainly acidic residues (aspartates), whereas the walls of the groove are lined with basic lysine and arginine residues. The electrostatic surface (Fig. 1C) calculated at pH 4 clearly shows a positively charged active-site groove required for interaction with the negatively charged polygalacturonan backbone.

From the alignment shown in Fig. S1, Asp207, Asp228 and Asp229 were identified as the catalytic residues of XghA. They are located on T3 turns 4 and 5 at the bottom of the groove. Based on homology with other GH28 enzymes, Asp228 acts as the proton donor (general acid), while either Asp207 and Asp229 (or both) are expected to activate the catalytic water molecule (base) (van Santen et al., 1999).

Other absolutely conserved residues in PGs are Asn205 (178 in PG2) and Tyr322 (291 in PG2), which are necessary for substrate binding (Shimizu et al., 2002). His251 (223 in PG2) maintains the proper ionization state of the general acid Asp228 (Armand et al., 2000). Lys286 (258 in PG2) and the cis-configuration of the peptide between Gly256 and Ser257 are necessary for binding of the subsite -1 C6 carboxylate of the substrate (Shimizu et al., 2002). The additional charges provided by the Nη atoms of Lys234 (Asn207 in PG2) and Lys281 (Asn253) (Fig. S1) may be important for the precise positioning of the C6 carboxylates of the xylosylated polygalacturonate rather than for interacting with a neutral xylose residue.

The four polygalacturonase sequence motifs, NTD (NTD205-207 in XghA), G/QDD (DDD 227-229 in XghA), G/SHG (SHG 250-253 in XghA), and RIK (GIK 284-286 in XghA) (Markovič et al., 2001), are largely conserved in XghA. The R→G substitution in the RIK motif is typical for xylogalacturonases. The loss of the positively charged Arg284 is compensated for by the Nη atom of Lys281. The smaller lysine compared to arginine may allow room for the xylose residue bound to the subsite -1 GalA residue.
A closer look at the active site showed the presence of two sulfate ions in the electron density map (Fig. 3). The first sulfate ion is bound at subsite -1 between the N and Oγ atoms of Ser257 of the cis-peptide bond, the OH of Tyr322, the Nη atoms of Lys234 and Lys286, and two water molecules. The second sulfate ion, at a distance of 6.4 Å from the first one, is bound at subsite +1 between Nδ of His251, OH of Tyr322, Nη of Lys286 and seven water molecules. This sulfate has an identical position to a sulfate ion observed in PG-1 (van Pouderoyen et al., 2003). Sulfate ions often mimic the positions of saccharide carboxylate groups (van Pouderoyen et al., 2003, van Asselt et al., 1999). Indeed, the two sulfate ions match the positions of carboxylates of the substrate bound to endo-PG1 from Stereum purpureum (Shimizu et al., 2002). Likewise, in Yersinia enterocolitica exopoligalacturonase the carboxylate of the bound GalA at subsite -1 has exactly the same position as the sulfate bound at subsite -1 in XghA (Abbott et al., 2007). From this, we conclude that Tyr322, Lys234, and Lys286 are probably involved in binding the substrate’s carboxylate at subsite -1, and His251, Tyr322, and Lys286 probably interact with the carboxylate of the sugar residue bound at subsite +1.

**Figure 3.** Close-up view of the active site of XghA. Sulfate ions, binding at the -1 and +1 subsites are shown in ball-and-stick representation. Catalytic residues, D207, D228 and D229, are shown. No interactions occur between the catalytic residues and the sulfate ions. Hydrogen bonds are indicated with dashes.

**Substrate modeling and molecular docking**

The major difference in the substrates of endo-PGs and XghA is the presence of xylose bound at the O3 of GalA. The conformation of GalA is restricted by the di-axial α-1,4-glycosidic bond and a weak O2···O6’ hydrogen bond (Braccini et al., 1999). In contrast,
the xyloses in XGA have more conformational flexibility due to their di-equatorial β-1,3-glycosidic linkage.

From the models of (XGA)$_2$ bound to XghA produced by AutoDock Vina only those were considered that had their galacturonic acid carboxylates at similar positions to the two sulfates in the active site in the crystal structure. Furthermore, candidate models had to have their reducing end (C1) directed toward the C-terminus and their non-reducing end (C4) directed toward the N-terminus of the protein, and both ends had to be accessible from the solvent. As the active site in XghA is open, many XGA models generated by AutoDock did not interact with the protein. The (XGA)$_2$ model that best fitted the criteria could be extended to the +2 subsite (Fig. 4), but extending this (XGA)$_3$ model to the -2 subsite only gave unacceptable models, probably because of insufficient interactions with the protein.

In subsite -1, the carboxylate has interactions with Lys234 and the backbone amide and hydroxyl of Ser257 (cis-peptide). There is enough space to accommodate a xylose residue attached to the O3 of GalA. This xylose residue may interact with the polar Asn174; in the PGs, the equivalent residue is a small hydrophobic residue. At position 175, a Val is present, which is smaller than the Gln conserved in endo-PGs, providing space for the xylose; a Gln at this position would probably clash with the xylose, thereby preventing binding and degradation of the substrate.

Figure 4. Stereoview of (XGA)$_3$ modeled in the active site of XghA. For clarity reasons only amino acids interacting with the substrate are shown at the +1 and -1 subsites. Catalytic residues, D207, D228 and D229, are not shown. Carbons of the GalA backbone are colored green and carbons of the xylose residues are colored cyan.
In subsite +1, the carboxylate shows interactions with His251 and Lys286. Sufficient space is also available for a xylose branch in this subsite. In contrast, in endo-PGs, the xylose would likely clash with His177 (equivalent to the more flexible Lys204) and with Asn126, which restricts access to the active site because of earlier unwinding of helix α3 in endo-PG. At subsite +2, the carboxylate of GalA interacts with Lys204; the xylose attached to the GalA in this subsite is bound by Thr280, Lys281, Gln317 and Gln319.

The scissile glycosidic bond between the -1 and +1 subsites is somewhat strained ($\varphi/\psi = -22/113$) compared to the ideal torsion angle values of polygalacturonan helices ($3_1$ helix, $\varphi/\psi = 80/89$; $2_1$ helix, $\varphi/\psi = 80/161$ (Scavetta et al., 1999) suggesting that interactions with the protein distort its low-energy conformation. In contrast, the glycosidic torsion angles of the β-1,3-bound xylose have more freedom (Table 2), and probably depend on precise interactions of the xylose with the protein. From the modeling it, may be concluded that the most extensive interactions of XghA with xylose exist at the +2 subsite, but no clear pocket is present. Degradation studies with GalA$_4$Xyl$_3$ as a substrate (in which the reducing end galacturonic acid is not xylosylated) (Zandleven et al., 2005) indicate fast formation of GalAXyl and GalA$_3$Xyl$_2$ products by XghA. Further conversion of the latter product to GalA$_2$Xyl was much slower. As XghA mainly acts from the non-reducing end on these small substrates, it can be concluded that a xylose residue at the +2 site increases the reaction rate significantly, in accordance with the crystal structure and the modeling study. The inability of XghA to degrade linear oligosaccharides of GalA or minimally β-xylose substituted GalA may be attributed to lack of interaction of the substrate with the subsites -1, +1 and +2.

**Table 2.** Torsion angles ($\varphi$ and $\psi$) of the glycosidic bonds in modeled XghA-(XGA)$_3$ complex structures.

Definitions of the glycosidic bond torsion angles between two (1,4)-linked GalA residues: $\varphi$ (phi) = ($O5_i - C1_i - O4_{i+1} - C4_{i+1}$) and $\psi$ (psi) = ($C1_i - O4_{i+1} - C4_{i+1} - C5_{i+1}$). Definitions of the glycosidic bond torsion angles between GalA and a (1,3)-linked Xyl: $\varphi$ (phi) = ($O5_i - C1_i - O3_{i+1} - C3_{i+1}$) and $\psi$ (psi) = ($C1_i - O1_{i+1} - C3_{i+1} - C4_{i+1}$).

<table>
<thead>
<tr>
<th>Glycosidic Bond</th>
<th>$\varphi$ (1,4)</th>
<th>$\psi$ (1,4)</th>
<th>$\varphi$ (1,3)</th>
<th>$\psi$ (1,3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-1,4</td>
<td>-22</td>
<td>113</td>
<td>60</td>
<td>84</td>
</tr>
<tr>
<td>+1/+2</td>
<td>105</td>
<td>66</td>
<td>124</td>
<td>12</td>
</tr>
<tr>
<td>$\beta$-1,3</td>
<td>74</td>
<td>99</td>
<td>74</td>
<td>99</td>
</tr>
</tbody>
</table>

XghA has been reported to be a processive, endo-acting enzyme (Zandleven et al., 2005, Beldman et al., 2003). Indeed, the extended open groove of XghA, where the xylogalacturon polymer binds, is in agreement with this endo-activity. Although an extended substrate-binding groove is a requirement for processivity (Breyer et al.,
2001), the structural basis for processivity is not obvious, as the substrate is only slightly enclosed by the walls of the groove. However, the enzyme binds the substrate at least at four adjacent sites at least (Zandl even et al., 2005), which may facilitate sliding of the substrate through the groove (Breyer et al., 2001), and thus be a determinant of processivity. In this respect, it is important to note that no obvious binding pockets for the xylose side chains exist that may hinder easy sliding motions.

The crystal structure of XghA constitutes the first endo-xylogalacturonan hydrolase 3D-structure described so far. The structure of XghA, although similar to polygalacturonase structures, reveals a much wider active-site groove and specific xylose-binding interactions at the +2 site. The -1 and +1 subsites also provide extra room for xylose residues. A substrate docking analysis explains the enzyme’s unique substrate specificity for xylosylated polygalacturonan. Further mutational experiments are required for validation of the assigned subsites for XghA binding.

**EXPERIMENTAL PROCEDURES**

**Protein expression and purification**

The enzyme xylogalacturonan hydrolase (XghA) from *Aspergillus tubingensis* was cloned (van der Vlugt-Bergmans et al., 2000) and expressed in *Aspergillus niger* PlugBug™ (van Dijck, 1999) (kindly provided by DSM Food Specialties, Delft, the Netherlands). The crude enzyme preparation was extensively dialyzed against 10 mM Tris buffer, pH 7.5, and filtered over a 0.22 μm filter to remove insoluble debris. XghA was partially purified by anion exchange chromatography on Q-Sepharose HP (GE Healthcare) with 20 mM Tris buffer, pH 7.5, and elution with 300 mM NaCl. The enzyme was further purified by gel filtration using a Superdex75 HR10/30 column (GE Healthcare, Diegem, Belgium), equilibrated with 20 mM Tris buffer, pH 7.5, containing 150 mM NaCl. XghA eluted at a molecular mass of 56 kDa; on SDS-PAGE the apparent molecular mass was ~ 55 kDa. XghA fractions were pooled and concentrated to 4.5 mg mL⁻¹ using an Amicon UltraCel-30K filter unit (Merck Millipore, Billerica, MA, USA). Dynamic light scattering experiments were performed using a DynaPro MS800TC instrument (Wyatt Technology Corporation, Santa Barbara, CA, USA) at 20 °C. Dynamic light scattering data were processed and analyzed with Dynamics (Wyatt Technology Corporation) software. From the dynamic light scattering experiments an apparent molecular mass of 30 kDa was deduced with a polydispersity of 20%.
Crystallization, data collection and structure determination

Initial vapour-diffusion crystallization experiments were performed using a Mosquito crystallization robot (Molecular Dimensions Ltd., Newmarket, UK). In a typical experiment, 0.1 μl screening solution was added to 0.1 μl protein solution on a 96-well MRC2 plate (Hampton Research, Aliso Viejo, CA, USA); reservoir wells contained 60 μl screening solution. The screening solutions used for the experiments were Structure Screen (Molecular Dimensions Ltd.), Wizard and Cryo (both from Emerald Biosystems, Bainbridge Island, WA, USA), and JCSG+ (Qiagen Systems, Germantown, MD). XghA crystals were obtained at room temperature from high ammonium sulfate concentrations. Optimization of the crystallization conditions yielded well-diffracting crystals that grew within several weeks to months when 2 μl protein solution (4.7 mg mL\(^{-1}\)) was mixed with 2 μl reservoir solution containing 2.9 M ammonium sulfate with 100 mM sodium acetate buffer, pH 4.4, and 15% glycerol.

A native dataset to 1.75 Å resolution was collected on beamline ID23-1 at the European Synchrotron Radiation Facility (Grenoble, France) at 100 K. Intensity data were processed using iMOSFLM (Battye et al., 2011) and the CCP4 package (Winn et al., 2011). The space group was I222, with unit cell dimensions of \(a = 76.4\), \(b = 121.2\), \(c = 129.7\) Å. With one monomer of 40 kDa in the asymmetric unit, the \(V_M\) is 3.8 Å\(^3\)/Da (Matthews, 1968) with a calculated solvent content of 67%. A summary of data collection statistics is given in Table 1.

Using the FFAS03 server (Jaroszewski, L., Rychlewski, L., Li, Z., Li, W. & Godzik, A., 2005) and SCWRL (Canutescu et al., 2003), homology models for XghA were generated. Four structures with highest identity were used as templates - \(A. niger\) Endo-PG I (PDB code 1NHC; (van Poudereyven et al., 2003)), \(A. aculeatus\) RGase A (PDB code 1RMG; (Petersen et al., 1997)), \(A. niger\) Endo-PG II (PDB code 1CZF; (van Santen et al., 1999)), and \(Colletotrichum lupini\) Endo-PG1 (PDB code 2IQ7; (Bonivento et al., 2008)) – with sequence identities of 27%, 25%, 22% and 22%, respectively. Molecular replacement was performed using PHASER (McCoy et al., 2007). ARP/wARP (Langer et al., 2008) was used for automatic building and the model was refined using REFMAC5 (Murshudov et al., 1997). Coot (Emsley et al., 2004) was used for manual rebuilding and map inspection. In 2Fo-Fc maps clear electron density was present for N-glycosylation at Asn278 and Asn301, and O-glycosylation at Thr38. The quality of the model was analyzed using MolProbity (Davis et al., 2007); secondary-structure elements were assigned with DSSP (Kabsch et al., 1983). Figures were prepared using PyMOL version 1.2r1 (Schrödinger LLC, New York, NY, USA) (DeLano, 2002) and ESPript (Gouet et al., 1999). Electrostatic profiles were calculated using the APBS (Adaptive Poisson-Boltzmann solver) tool interfaced to PyMOL after a PDB2PQR conversion (Andre-Leroux et al., 2009). Calculations were performed at pH 4 using the AMBER force field. Atomic
coordinates and experimental structure factor amplitudes for XghA have been deposited in the RCSB Protein Data Bank under accession code 4C2L.

Substrate modeling and molecular docking

As a starting model for molecular docking, the experimentally observed digalacturonate bound to Yersinia enterocolitica YeGH28 (PDB entry 2UVF, (Abbott et al., 2007)) was used. Xylose residues were added to the C3 atoms of the two galacturonic acids using the PRODRG server (Schüttelkopf et al., 2004), resulting in the substrate (XGA)$_2$. The modeled substrate was used for automated docking to XghA using AutoDock Vina (Trott et al., 2010) with the aid of the PyMOL plug-in (Seeliger et al., 2010). All solvent molecules were removed from the XghA structure except for the catalytic water molecule. XghA and (XGA)$_2$ were treated as a rigid protein and a flexible molecule, respectively. The grid of the docking simulation was defined by a 22 $\times$ 36 $\times$ 25 Å box centered on the XghA -1 and +1 subsites. The docking simulation was performed using the default parameters. The 20 top-ranked generations based on the predicted binding affinity (kcal.mol$^{-1}$) were analyzed. The best docked (XGA)$_2$ molecule was extended to the +2 subsite by adding a XGA moiety to (XGA)$_2$. The resulting (XGA)$_3$ molecule was used for another docking simulation.

ACKNOWLEDGEMENTS

We thank Marga Herweijer (DSM Food Specialties, Delft, the Netherlands) for her kind preparation of XghA. We are grateful to the staff scientists at beam line ID23-1 at the ESRF, Grenoble, France, for their support during data collection.

SUPPORTING MATERIAL

Figure S1. Structure–based alignment of GH28 enzymes.
Structure–based alignment of XghA from Aspergillus tubingensis, endo-polygalacturonase I from Aspergillus niger (PG1), endo-polygalacturonase II from Aspergillus niger (PG2), polygalacturonase from Aspergillus aculeatus (PGA), and rhamnogalacturonase A from Aspergillus aculeatus (RGase). The structural alignment was made with Dali (Holm et al., 2000). The secondary structure elements above the sequence alignment are those obtained from the crystal structure of XghA. Identical residues have a red background color, similar residues have a red color. Residues in the -1 subsite are shown with a purple background, and the catalytic residues have a cyan background. Residues marked with a * (asterisk) have double conformations. Disulfide linkages are indicated in green italics below the sequences. The figure was created with ESPript (Gouet et al., 1999).