BJ www.biochemj.org

BJ Structure

Biochem. J. (2009) **418**, 39–47 (Printed in Great Britain) doi:10.1042/BJ20081256

Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family

Elien VANDERMARLIERE*, Tine M. BOURGOIS†, Martyn D. WINN‡, Steven VAN CAMPENHOUT†, Guido VOLCKAERT†, Jan A. DELCOUR§, Sergei V. STRELKOV*, Anja RABIJNS* and Christophe M. COURTIN§[1]

*Laboratory for Biocrystallography, Department of Pharmaceutical Sciences, Katholieke Universiteit Leuven, Herestraat 49, O&N II, bus 822, 3000 Leuven, Belgium, †Laboratory of Gene Technology, Department of Biosystems, Katholieke Universiteit Leuven, Kasteelpark Arenberg 21, bus 2462, 3001 Leuven, Belgium, ‡Computational Science and Engineering Department, STFC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, U.K., and §Laboratory of Food Chemistry and Biochemistry, Department of Microbial and Molecular Systems, Katholieke Universiteit Leuven, Kasteelpark Arenberg 20, bus 2463, 3001 Leuven, Belgium

AXHs (arabinoxylan arabinofuranohydrolases) are α-L-arabinofuranosidases that specifically hydrolyse the glycosidic bond between arabinofuranosyl substituents and xylopyranosyl backbone residues of arabinoxylan. *Bacillus subtilis* was recently shown to produce an AXH that cleaves arabinose units from O-2- or O-3-mono-substituted xylose residues: *Bs*AXH-m2,3 (*B. subtilis* AXH-m2,3). Crystallographic analysis reveals a two-domain structure for this enzyme: a catalytic domain displaying a five-bladed β-propeller fold characteristic of GH (glycoside hydrolase) family 43 and a CBM (carbohydrate-binding module) with a β-sandwich fold belonging to CBM family 6. Binding of substrate to *Bs*AXH-m2,3 is largely based on hydrophobic stacking interactions, which probably allow the positional flexibility needed to hydrolyse both arabinose substituents at the O-2 or O-3 position of the xylose unit. Superposition of the *Bs*AXH-m2,3 structure with known structures of the GH family 43 exo-acting enzymes, β-xylosidase and α-L-arabinanase, each in complex with their substrate, reveals a different orientation of the sugar backbone.

Key words: arabinoxylan arabinofuranohydrolase, *Bacillus subtilis*, crystallography, enzyme–substrate complex, family 43 glycoside hydrolase, substrate-binding mechanism.

INTRODUCTION

The use of renewable sources as an alternative to petroleum-derived fuels and materials is receiving more and more research attention. At the moment, the most common renewable fuel is ethanol derived from cereal starch. Because of limitations in cereal supply, a better utilization of the whole kernel is being explored, with a focus on hemicellulose [1,2]. An important hemicellulose is arabinoxylan, which is composed of a linear backbone of β-1,4-linked D-xylopyranosyl units that can be O-2 and/or O-3 substituted with L-arabinofuranosyl units. Enzymatic degradation of arabinoxylan into its building blocks requires a set of xylanolytic enzymes. Endoxylanases and β-xylosidases hydrolyse the xylan backbone and act in synergy with arabinofuranosidases, which remove the arabinofuranosyl units [3]. These enzymes are thus important players in the conversion of heteroxylan into xylose and arabinose, which can then be used either for the production of ethanol or as platform molecules.

α-L-Arabinofuranosidases (EC 3.2.1.55) are classified according to their substrate specificity. Type A arabinofuranosidases are only active against small substrates such as 4-nitrophenyl α-L-arabinofuranoside and short-chain AXOSs (arabinoxylo-oligosaccharides). Type B arabinofuranosidases can also hydrolyse polymeric substrates such as branched arabinan and arabinoxylan [4]. Some type B arabinofuranosidases specifically cleave arabinofuranosyl units from arabinoxylan and hence are termed AXH (arabinoxylan arabinofuranohydrolase) [5]. The latter type can be further divided into two and possibly three groups. AXH-m releases arabinose only from mono-substituted xylose units, whereas AXH-d releases it only from double-substituted xylose units [6,7]. Ferré et al. [8] suggested that AXH from barley malt releases arabinose from both single- and double-substituted xylose units and therefore can be classified as AXH-md.

Based on sequence similarity, the currently known arabinofuranosidases have been classified into five GH (glycoside hydrolase) families. Those found in GH families 3, 51 and 54 hydrolyse the glycosidic bond with retention of the anomeric configuration. Those in GH families 43 and 62 invert the anomeric configuration on hydrolysis [9,10]. To date, the only known structures of arabinofuranosidases are those of arabinofuranosidases belonging to GH families 51 and 54, displaying a $(\beta/\alpha)_8$-barrel and a β-sandwich respectively [11,12].

Recently, Bourgois et al. [13] have characterized XynD from *Bacillus subtilis* subspecies *subtilis* A.T.C.C. 6051, which was previously predicted to be a member of GH family 43 displaying endoxylanase activity supplemented with arabinofuranosidase co-activity. Careful biochemical analysis characterized it as an arabinoxylan arabinofuranohydrolase that cleaves arabinose units from O-2- or O-3-mono-substituted xylose residues, i.e. as an AXH-m2,3 [13]. Within GH family 43, structures of β-xylosidases from *Geobacillus stearothermophilus* [14], *Bacillus halodurans*, *B. subtilis* and *Clostridium acetobutylicum* are available. Also α-L-arabinanases from *Cellvibrio japonicus* [15] and *Bacillus thermodenitrificans* [16] and a bifunctional β-xylosidase/α-L-arabinofuranosidase from *Selenomonas ruminantium* [17] belonging to GH family 43 had their structure determined. Together with GH families 32, 62 and 68, their catalytic

Abbreviations used: AXH, arabinoxylan arabinofuranohydrolase; AXOS, arabinoxylo-oligosaccharide; *Bs*AXH-m2,3, *Bacillus subtilis* AXH-m2,3; CBM, carbohydrate-binding module; GH, glycoside hydrolase.

[1] To whom correspondence should be addressed (email Christophe.Courtin@biw.kuleuven.be).

domain shares a five-bladed β-propeller fold which was first seen for tachylectin [18]. Unlike the classical organization of the active site of inverting enzymes, which consists of a proton donor (the general acid) and a nucleophile (the general base), a third carboxylate was found to be essential for activity in GH family 43 enzymes. This residue is responsible for the pK_a modulation of the general acid and correct orientation of both the proton donor and substrate [14]. Although the resolved α-L-arabinanase structures show no associated CBM (carbohydrate-binding module), other GH family 43 members contain a CBM. For *Bs*AXH-m2,3 (*B. subtilis* AXH-m2,3), it has been proposed, based on sequence similarity, that this CBM belongs to CBM family 6 [13]. Members of this family display a β-sandwich fold: a lectin-like β-jelly roll consisting of five antiparallel β-strands packed against four antiparallel β-strands. Like other CBMs, their function is to increase the effective concentration of the active site on polymeric substrate [19,20].

Since β-xylosidases, α-L-arabinanases and bifunctional β-xylosidases/α-L-arabinofuranosidases also belong to GH family 43, they not only have their structure in common, but also an analogous reaction mechanism as their catalytic residues occur at an analogous position [10]. However, unlike β-xylosidases and α-L-arabinanases, which are exo-acting enzymes, AXH-m2,3 releases the arabinose substituents of arabinoxylan, so we can expect differences in the substrate-binding mechanism. To understand the binding of substrate, knowledge of the three-dimensional structure is essential. For β-xylosidase and α-L-arabinanase, the structure in complex with xylobiose [14] and the structure in complex with arabinohexaose [15] respectively are available and will be used here as reference structures. No structure of an AXH-m2,3 is available. We have previously reported the crystallization and preliminary X-ray analysis of *Bs*AXH-m2,3 in complex with xylotriose [21]. Here, we describe the crystal structure of *Bs*AXH-m2,3, the first structure of a GH family 43 arabinofuranosidase, and the first structure of an AXH overall, in complex with several substrates. The structure allows in-depth comparison of the substrate-binding mechanism between the GH family 43 classes.

EXPERIMENTAL

Crystallization and data collection

Recombinant AXH-m2,3 from *B. subtilis* was expressed and purified as described by Bourgois et al. [13]. Crystallization and data collection of *Bs*AXH-m2,3 crystals soaked in xylotriose were reported earlier [21]. In short, thick needle-like crystals were grown in 4.0 M sodium formate. Refinement of this condition using the Additive screen (Hampton Research) resulted in several conditions in which rod-like crystals could be grown (Table 1). All these different crystals belong to the same space group and were used at random to prepare soaks with several sugars. Prior to data collection, the crystals were transferred briefly into cryoprotectant composed of 4.0 M sodium formate supplemented with 30 % (v/v) glycerol and a saturated concentration of xylotriose (Megazyme, Bray, Ireland) [21]. Analogous soaking experiments were performed using xylotetraose (Megazyme), cellotetraose (Sigma–Aldrich, St. Louis, MO, U.S.A.) and a mixture of AXOSs with an average degree of polymerization of 4 and an arabinose-to-xylose ratio of 0.5 (AXOS-4-0.5). These AXOSs were derived from WPC (wheat pentosan concentrate; Pfeifer & Langen, Dormagen, Germany) and were kindly made available by Katrien Swennen (Laboratory of Food Chemistry and Biochemistry, K.U. Leuven). In all cases, a saturated solution was used and soaking times ranged from 1 to 5 min.

Data from *Bs*AXH-m2,3 crystals soaked in xylotriose and xylotetraose were collected on beamline BW7a of DESY (EMBL-Hamburg, Hamburg, Germany) at cryogenic temperature and processed using the HKL suite of programs [22]. Datasets of native *Bs*AXH-m2,3 crystals and those soaked in cellotetraose and AXOS-4-0.5 were collected on beamline X10SA of SLS (Swiss Light Source; Villigen, Switzerland). These data were processed with iMosflm [23,24] and scaled and merged with Scala [25]. For all structures, 5 % of the observations were set aside for cross-validation analysis. These sets are all equal to the set of observations of the data collected of the *Bs*AXH-m2,3 crystal soaked in xylotriose.

Structure solution and refinement

All further computing was done using the CCP4 suite [26] unless stated otherwise. Due to the relatively low sequence similarity {25 % when comparing the catalytic domain with that of *Ce. japonicus* α-L-arabinanase (PDB entry 1gyh) [15] and 35 % when comparing the CBM with *B. halodurans* CBM family 6 (PDB entry 1w9t) [27]}, phasing the structure of *Bs*AXH-m2,3 by molecular replacement presented a challenging problem. Eventually, the structure of *Bs*AXH-m2,3 soaked in xylotriose was solved with an automatic approach: MrBUMP [28] implemented in the CCP4 interface, and using Phaser [29] as the underlying molecular replacement program, was used to position models of the two domains. An initial attempt at automated model rebuilding with ARP/wARP [30] using the sequence of *Bs*AXH-m2,3 failed. Phase improvement of the molecular replacement solution was then carried out using artificial phase extension and dynamic density modification as implemented in the latest version of ACORN [31]. This step improved the phases sufficiently for ARP/wARP [30] to build and dock 93 % of the residues. The co-ordinates were evaluated using Coot [32] and few corrections were made manually. The refined structure was then used to solve the structure of the native protein and the remaining substrate complexes using Phaser. After several cycles of refinement, residual positive density in the $F_o - F_c$ and $2F_o - F_c$ electron density revealed the presence of several bound sugar molecules. These were inserted manually followed by further refinement and model building using Refmac5 [33] and Coot respectively. The final structures were evaluated using Molprobity [34]. All data collection and refinement statistics are shown in Table 1. Figures 1–5 were drawn using the program PyMOL (DeLane Scientific; http://pymol.sourceforge.net/).

RESULTS AND DISCUSSION

The overall structure

The structure of AXH-m2,3 from *B. subtilis* was determined by molecular replacement using the *Ce. japonicus* α-L-arabinanase (PDB entry 1gyh) [15] and *B. halodurans* CBM family 6 (PDB entry 1w9t) [27] structures as search models for the catalytic domain and CBM respectively. Several soaking experiments were performed to gain insight into the binding of substrate. All structures were determined to a maximum resolution ranging between 1.55 and 2.05 Å (1 Å = 0.1 nm) and their refinement converged to good *R*-factors. This indicates that the structures can be used for a reliable analysis of the different interactions. Inspection of the electron density maps of the soaked structures revealed clear density for several sugar units. For the soaking experiments with xylotriose and xylotetraose respectively, three and four xylose units could be built in the electron density and the bound ligand atoms refined well with average thermal factors

Table 1 Data collection and refinement statistics

Values in parentheses are for the highest resolution shell.

	AXH	AXH xylotriose	AXH xylotetraose	AXH AXOS-4-0.5	AXH cellotetraose
Additive	30 % (w/v) Sucrose	1.0 M Lithium chloride		0.1 M Betaine hydrochloride	30 % Sucrose
Data collection					
Wavelength (Å)	0.9785	1.0788	1.0322	0.9785	1.0011
Source	SLS X10SA	DESY BW7a	DESY BW7a	SLS X10SA	SLS X10SA
Resolution range (Å)	30–2.0 (2.11–2.0)	50–1.55 (1.58–1.55)	50–2.05 (2.09–2.05)	30–1.8 (1.9–1.8)	30–1.8 (1.9–1.8)
Reflections					
Observed	156621	258631	150581	390182	306957
Unique	35773 (4966)	76939 (3868)	33537 (1596)	50933 (7349)	45414 (6055)
Completeness (%)	96.4 (93.2)	98.1 (99.9)	97.0 (93.9)	99.9 (99.8)	92.0 (85.4)
Mean $I/\sigma(I)$	14.1 (3.9)	14.0 (3.9)	22.4 (8.0)	19.9 (6.1)	19.0 (5.4)
Multiplicity	4.4 (4.2)	3.4 (3.1)	4.5 (4.2)	7.7 (7.5)	6.8 (6.6)
R_{sym} (%)*	10.2 (33.3)	8.3 (32.2)	4.6 (13.7)	8.1 (27.9)	8.6 (29.2)
Crystal					
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters	$a = 67.6$ Å	$a = 68.7$ Å	$a = 67.5$ Å	$a = 67.8$ Å	$a = 68.0$ Å
	$b = 74.4$ Å	$b = 73.7$ Å	$b = 72.4$ Å	$b = 74.2$ Å	$b = 73.1$ Å
	$c = 107.0$ Å	$c = 106.5$ Å	$c = 106.9$ Å	$c = 107.3$ Å	$c = 105.7$ Å
Refinement					
R_{work} (%)†	16.8	16.4	15.1	15.5	15.1
R_{free} (%)‡	21.3	19.0	17.4	19.0	17.5
Root mean square deviations					
Bond lengths (Å)	0.016	0.010	0.012	0.015	0.012
Bond angles (°)	1.50	1.22	1.50	1.44	1.50
Number of atoms					
Protein	3671	3674	3675	3682	3661
Solvent	311	579	446	407	440
Ligand		28	37	28	45
Average *B*-factor (Å^2)					
Main chain	16.1	13.9	12.5	14.1	14.9
Side chain	16.0	14.5	12.5	14.1	15.4
Solvent	24.4	29.1	24.0	24.1	32.0
Ligand		16.7	18.1	52.3	60.0
PDB entry	3C7E	3C7F	3C7G	3C7H	3C7O

*$R_{sym} = \Sigma|I(I)|/\Sigma I$.
†$R = \Sigma||F_o|-|c|/\Sigma|F_c|$.
‡The R_{free} was calculated with 5 % of the data excluded from structure refinement.

comparable with the values for all protein atoms. Also for the crystals soaked in cellotetraose and AXOS-4-0.5, there was clear density for four glucose and four xylose units respectively, but no interpretable density for arabinose units could be observed for the structure soaked in AXOS-4-0.5. For these structures, the thermal factors of the sugars are slightly higher compared with those for all protein atoms (Table 1).

Each asymmetric unit contains one *Bs*AXH-m2,3 monomer, which is organized into two domains: the N-terminal catalytic domain, which is a five-bladed β-propeller fold common to all GH family 43 members, and a C-terminal β-sandwich domain, the CBM, analogous to CBM family 6 members (Figure 1). Superposition of the different structures revealed no large structural changes in the overall structure on ligand binding.

Unlike the β-xylosidase from *G. stearothermophilus* [14] and the bifunctional β-xylosidase/α-L-arabinofuranosidase from *S. ruminantium* [17], *Bs*AXH-m2,3 does not crystallize as a homotetramer, but as a single entity. This monomeric state of *Bs*AXH-m2,3 is in contrast with the quaternary structure of other two-domain GH family 43 members for which the structure is known. The latter all occur as homotetramers, which are composed of two dimers turned at 90° against each other. The monomers forming the dimer are aligned antiparallel to one another such that the catalytic domain of one monomer interacts with the CBM of the neighbouring domain. This arrangement results in a quaternary structure exhibiting 222 symmetry.

The catalytic domain

The catalytic domain forms a five-bladed β-propeller, with the five β-sheets (marked I–V) radially arranged around a central water-filled tunnel like the blades of a propeller. Each β-sheet is built up of four antiparallel β-strands connected by hairpin turns, with the first strand being the innermost. All five innermost β-strands start at the same site, i.e. the entrance site, where the active site is located, and run almost parallel to one another. β-Strands 2 and 3 of the five β-blades are connected by large loops lining the active site (Figure 1b). Together with the loop regions connecting β-strand 4 of one sheet to strand 1 of the next sheet, these loop regions probably have a role in determining the substrate specificity of *Bs*AXH-m2,3.

Unlike most propeller structures, *Bs*AXH-m2,3 does not show the classical molecular velcro, which closes the propeller by incorporating the N-terminus and C-terminus in one blade of the propeller. However, some hydrogen bonds exist between the N-terminus and strand 4 of propeller blade V. These bonds probably provide stabilization to the fold. This non-velcroed propeller

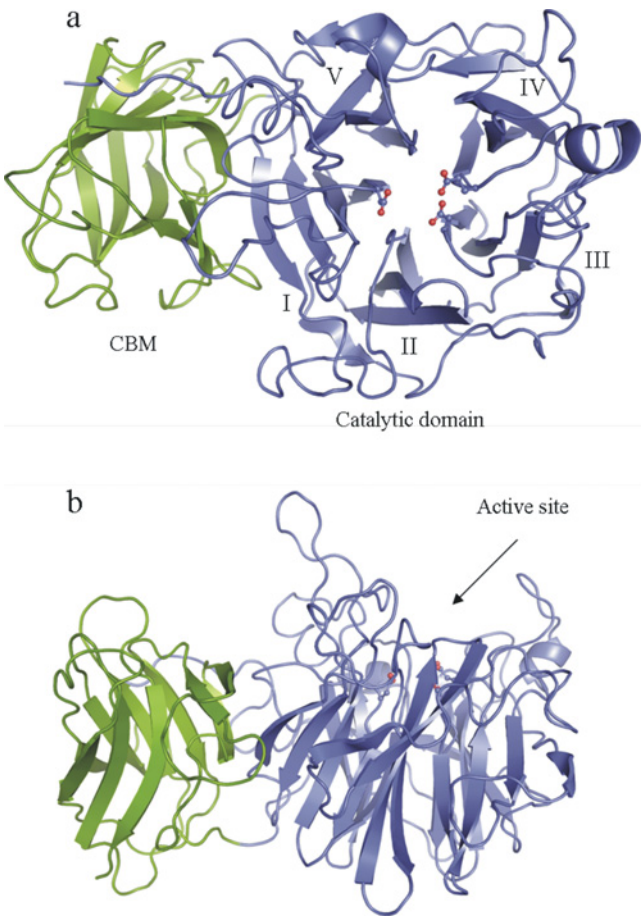

Figure 1 Overall structure of *Bs*AXH-m2,3

The catalytic domain is shown in blue, while the CBM is shown in green. The three catalytic residues are shown in ball and stick representation. (**a**) Top view showing the numbering of the five β-blades (I–V). (**b**) Side view showing the position of the active site surrounded by the long loops connecting β-strands 2 and 3 of each β-blade.

has also been observed in the structure of other known GH family 43 members [15].

The active site and substrate binding

For members of GH family 43, three residues are essential for catalytic activity. For *Bs*AXH-m2,3, these residues could be identified by superposing the structure of *Bs*AXH-m2,3 with the structure of the β-xylosidase from *G. stearothermophilus* [14]. They are: Asp-24, which is the general base; Glu-225, which is the general acid; and Asp-163, which is believed to play a role in the pK_a modulation of the general acid and correct orientation of both the general acid and substrate. They are located on the innermost strands of β-blades I, IV and III respectively and point to the centre of the tunnel entrance site. Seen from the surface, they are situated in a small pocket, in which a single arabinose substituent would fit well, along an open cleft, which allows binding of the xylan backbone. This arrangement of the active site is very suitable for the hydrolysis of arabinose substituents from mono-substituted xylose units and nicely illustrates why *Bs*AXH-m2,3 does not hydrolyse arabinose from disubstituted xylose units. The pocket is too small to allow binding of two arabinoses.

Since the crystals are grown from active *Bs*AXH-m2,3, soaking with its substrate, AXOSs, should result in hydrolysis of the arabinose substituents, and hence only the bound xylan backbone would probably be observed. Therefore soaking experiments were performed not only with AXOSs but also with short unsubstituted xylan chains to gain insight into the binding of the xylan backbone to *Bs*AXH-m2,3. Three-dimensional complexes were obtained with xylotriose, xylotetraose, AXOS-4-0.5 and cellotetraose. The latter soaking experiment was performed to gain insight into the binding capacity of the CBM to cellulose, since some CBM family 6 members are found to bind cellulose [19]. Surprisingly, cellotetraose was bound to the active site. To determine the direction of the sugar backbone, the difference electron density map was contoured at a high level to observe primarily the positions of the oxygen atoms. Interestingly, for the complex structures, a glycerol molecule (originating from the cryo solution) is located at the probable position of the target arabinose and hence gives us a hint about the interactions with the arabinose unit in this subsite. The different complex structures reveal several residues responsible for binding interactions with the xylan backbone. Since *Bs*AXH-m2,3 hydrolyses substituents of the xylan backbone, the subsite numbering proposed by Davies et al. [36] cannot be used to number the different binding subsites for the xylose units of the xylan backbone. So, for convenience, the binding subsites observed here are numbered I–IV starting from the xylose unit at the reducing end of the xylan backbone, with the III subsite being equal to the –1 subsite in the numbering proposed by Davies et al. (Figure 2a). An arabinose substituent would be situated at the +1 subsite.

Superposition of the different complex structures shows no difference in the position of the bound sugars and glycerol molecules (Figure 2b). When superposing the unbound structure with the complex structure, the side chain of Asn-288 turns by 90°, away from the sugar. Apart from the xylose unit in the III subsite (from where the substituent is removed), only a few hydrogen-bonding interactions are observed between *Bs*AXH-m2,3 and the xylan backbone, of which all the xylose units are in the chair conformation (Figure 2). Two pronounced hydrophobic stacking interactions are observed with Phe-244 and Trp-160 and the xylose units in the II and IV subsites respectively, while the xylose unit in subsite I makes several hydrogen bonds with Gly-286. In the III subsite the OH2 of the xylose unit makes one hydrogen bond with Asn-288 (3.3 Å) and two relatively strong hydrogen bonds with the general acid, Glu-225 (2.6 and 3.0 Å), while the OH3 makes one hydrogen bond with Glu-225 (3.6 Å). The extra carbonyl group of the glucose units from cellotetraose does not seem to make any additional interaction with *Bs*AXH-m2,3 in comparison with xylotetraose. The binding of cellotetraose to the active site is probably an artefact based on the structural relationship between the xylopyranose and glucopyranose ring, since this binding has no physiological meaning, i.e. arabinocellulose does not occur in Nature.

In the II and IV subsites, binding is solely based on hydrophobic stacking interactions. Since, in contrast with an extensive hydrogen-bonding network, hydrophobic stacking interactions allow more positional freedom, these interactions may account for the observation that either substituents at the O-2 or O-3 hydroxyl position of the xylose can be hydrolysed. This property of *Bs*AXH-m2,3 needs some positional flexibility to allow correct positioning of the glycosidic bond with regard to the catalytic triad. Further studies involving soaking experiments with inactive *Bs*AXH-m2,3 and specific AXOSs only substituted at their O-2 or O-3 hydroxyl position are required to confirm the above hypothesis.

Since the hydroxy groups from glycerol usually take similar positions to the hydroxy groups from sugars, the interactions between the glycerol in the catalytic pocket and *Bs*AXH-m2,3 give

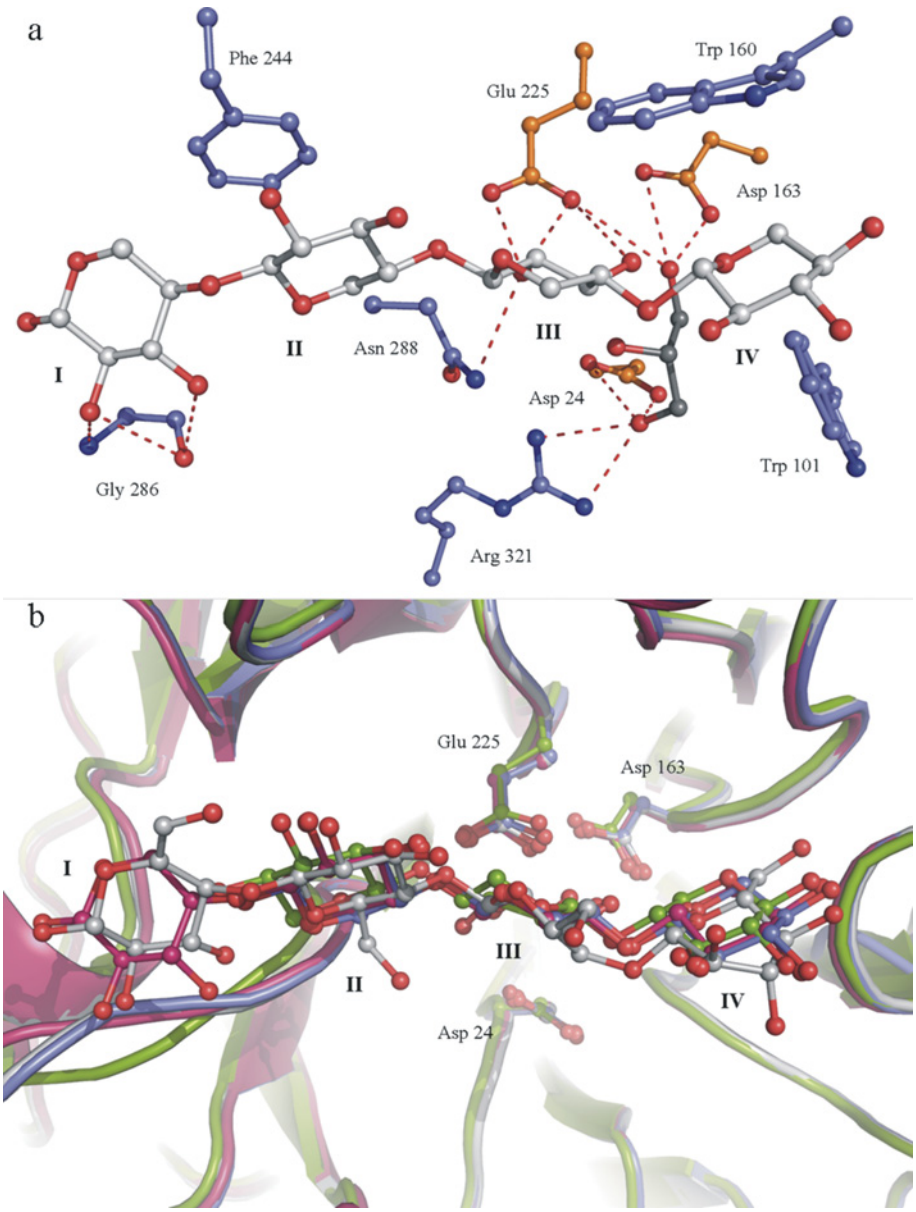

Figure 2 Detailed view of the active site of *Bs*AXH-m2,3

(**a**) The interactions between xylotetraose (white) and AXH-m2,3 are shown with hydrogen bonds represented as red dotted lines. The glycerol molecule occupying the catalytic pocket is shown in grey. The catalytic residues are in orange. (**b**) Superposition of the substrate molecules in the active site. The structure soaked in xylotriose is shown in blue, the one soaked in xylotetraose in pink and that with AXOS-4-0.5 in green, and the structure soaked in cellotetraose is shown in grey.

us an idea of the interactions with the arabinose unit in the catalytic pocket. Hydrogen bonds are formed with the three catalytic residues (Glu-225, Asp-163 and Asp-24) and Arg-321 (Table 2). In analogy with other GH family 43 members, Trp-101, which is part of the invariant Trp-Ala-Pro element, probably makes a hydrophobic stacking interaction with the arabinose unit [14].

The CBM

In analogy with other CBM family 6 members, the CBM of *Bs*AXH-m2,3 forms a β-sandwich fold consisting of five antiparallel β-strands on one face and four antiparallel β-strands on the other face (Figure 3). During refinement, the electron-density maps revealed residual density for metal ions at two positions (Figure 3). The first metal ion could be modelled as a calcium ion and is hepta-co-ordinated by the side chains of Glu-359, Glu-361, Asn-383 and Asp-480. The co-ordination is completed by the backbone carbonyl oxygen atoms of Gln-384 and Asp-480. The position and nature of this ion are identical with those found in other known CBM family 6 members and probably make it a structural ion [14,19,27,37,38]. It is likely to be already incorporated during protein folding, since no calcium ions were used during the purification and crystallization of *Bs*AXH-m2,3, supporting the idea of it being a structural ion. The co-ordination of the second ion, modelled as a sodium, involves the side chains of Gln-390 and Asp-393. This co-ordination is completed by the main-chain carbonyl oxygen atoms of Arg-368 and Ser-388 and two water molecules. This metal ion corresponds to the sodium ion observed in the CBM family 6 of *B. halodurans*

Table 2 Interactions between *Bs*AXH-m2,3 and xylotetraose

Subsite	Substrate atom	Protein atom	Distance (Å)
I subsite	O-2	Gly-286 N	3.2
	O-2	Gly-286 CO	3.5
	O-3	Gly-286 CO	3.6
II subsite		Phe-244	Hydrophobic stacking
III subsite (–1 subsite)	O-2	Asn-288 Nδ2	3.3
	O-2	Glu-225 Oε2	2.6
	O-2	Glu-225 Oε1	3.0
	O-3	Glu-225 Oε1	3.6
IV subsite		Trp-160	Hydrophobic stacking
+1 subsite (glycerol)	O-1	Glu-225 Oε1	3.4
	O-1	Asp-163 Oδ2	3.3
	O-1	Asp-163 Oδ1	2.5
	O-3	Asp 24 Oδ1	3.5
	O-3	Asp 24 Oδ2	2.7
	O-3	Arg-321 NH2	2.9
	O-3	Arg-321 NH1	2.8
		Trp-101	Hydrophobic stacking

laminarinase and to the sodium ion observed in the CBM family 6 of *Cellvibrio mixtus* endoglucanase 5A [27,38]. In contrast with the calcium ion, sodium might have been incorporated during the crystallization of *Bs*AXH-m2,3 since the precipitant and cryo solutions contain sodium ions.

Soaking experiments showed that neither cellotetraose nor xylo-oligosaccharides were bound to the CBM. This can tentatively be explained as follows. CBMs of family 6 contain two clefts that play a role in sugar binding (Figure 4a). Cleft A is found in the loop region connecting the inner and outer β-sheets of the β-sandwich and resembles the sugar-binding sites of lectins [37], whereas cleft B is located on the concave surface of one β-sheet and has a similar location to the binding sites of CBMs from several other families [38]. When comparing cleft A of *Bs*AXH-m2,3 with several other CBMs of family 6, *Bs*AXH-m2,3 is clearly lacking the two aromatic residues that play a key role in ligand binding [39]. The amino acids at the equivalent position are Ser-388 and Ala-445 (Figure 4b). Cleft B is lined with two aromatic residues, Trp-394 and Phe-465, which is in agreement with other CBMs belonging to family 6. However, in analogy with the CBM family 6 from *Clostridium thermocellum* xylanase 11A [19], the CBM family 6 from *B. halodurans* laminarinase [27] and the CBM family 6 from *Clostridium stercorarium* xylanase [37], a loop region covers this cleft, making it inaccessible for sugar (Figure 4c). These findings may explain the absence of binding of xylo-oligosaccharides and cellotetraose to the CBM of *Bs*AXH-m2,3, since both binding clefts are structurally incapable of binding sugar chains. This CBM probably lost its carbohydrate-binding function during evolution. The loss of carbohydrate-binding capacity is also seen for CBMs of family 48. This family of CBMs is related to the starch-binding domains belonging to CBM families 20 and 21. Members of the latter families have binding sites that are lined with aromatic residues [40]. In analogy with cleft A of the CBM of *Bs*AXH-m2,3, the carbohydrate-binding capacity of the CBM48 of glycosyltrehalose trehalohydrolase from *Sulfolobus solfataricus* is also lost due to mutations of the aromatic residues that are responsible for carbohydrate binding [41].

Comparison of *Bs*AXH-m2,3 with other GH family 43 structures

The availability of the structure of this AXH-m2,3 and of other GH family 43 enzymes, in complex with their substrates, provides insight into how these structurally related enzymes are fine-tuned towards different substrate specificities. Indeed, despite their similar structures, reaction mechanisms and catalytic residue

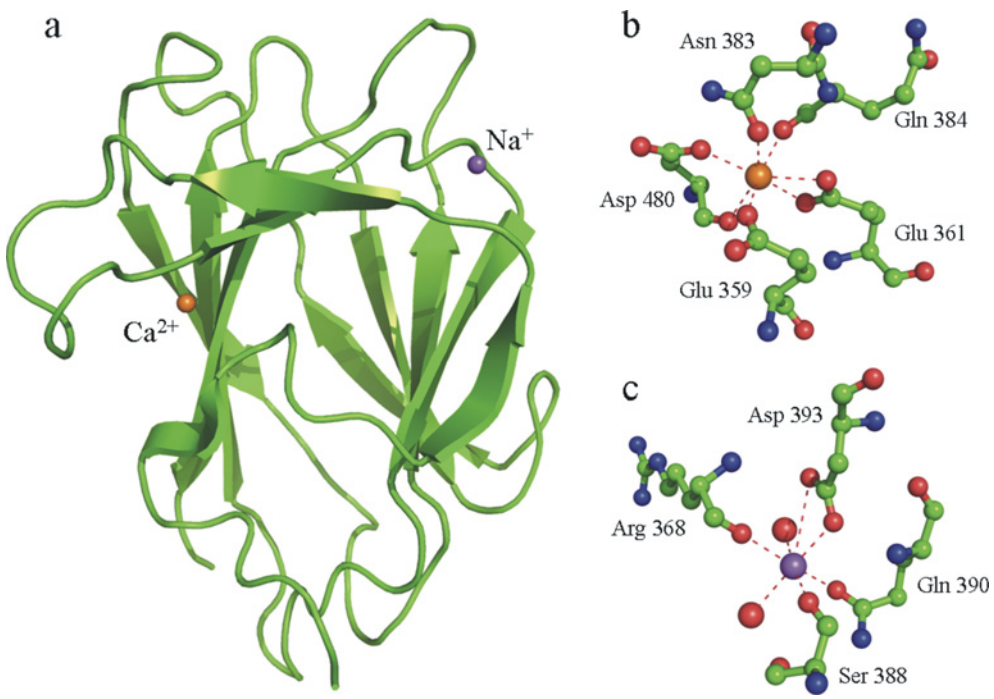

Figure 3 Overall structure of the CBM and its bound metal ions

(**a**) Overall view of the CBM. The sodium ion is shown in purple, the calcium ion in orange. (**b**) Detailed view of the interactions between *Bs*AXH-m2,3 and the calcium ion. (**c**) Detailed view of the interactions between *Bs*AXH-m2,3 and the sodium ion. The water molecules involved in co-ordinating the sodium ion are shown as red balls.

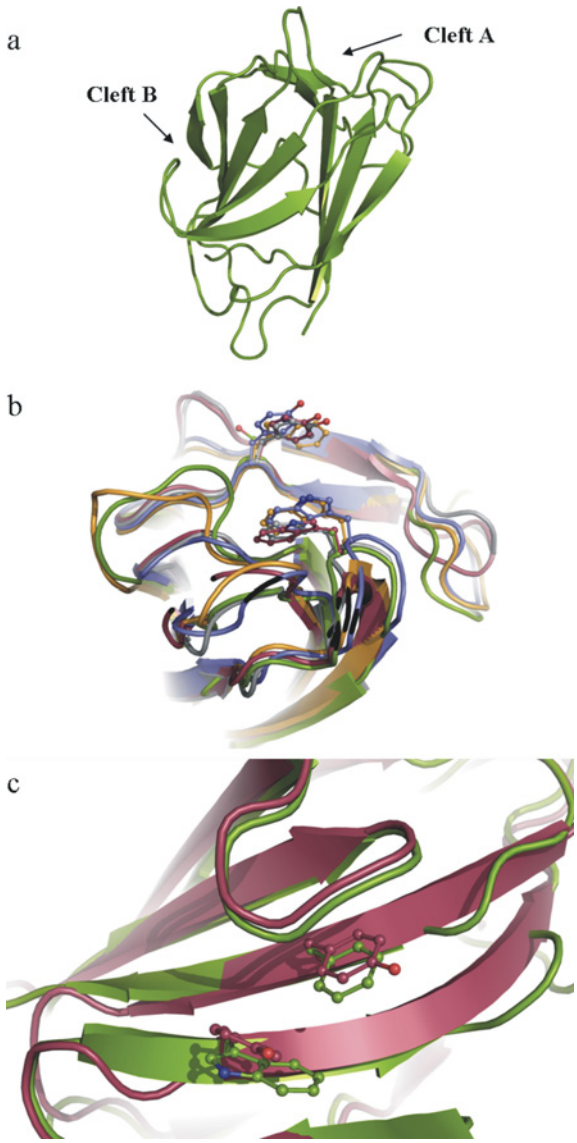

Figure 4 Structural differences at the binding clefts of CBM family 6 members

(**a**) Cartoon representation of the CBM of *Bs*AXH-m2,3 showing the position of both binding clefts A and B. (**b**) Detail of binding cleft A. The two aromatic residues with a key role in sugar binding are shown in ball and stick representation. The CBM6 of *Bs*AXH-m2,3 is shown in green, the CBM6 of *Cl. thermocellum* xylanase 11A (PDB entry 1gmm) [19] is shown in purple, the CBM6 of *Ce. mixtus* endoglucanase 5A (PDB entry 1uyz) [38] is shown in blue, the CBM6 of *Bacillus haludorans* laminarinase (PDB entry 1w9t) [27] is shown in orange and the CBM6 of *C. stercorarium* xylanase (PDB entry 1nae) [37] is shown in grey. (**c**) Detail of binding cleft B. The two aromatic residues that are important for sugar binding are shown in ball and stick representation. The CBM6 of *Bs*AXH-m2,3 is shown in green and the CBM6 of *Cl. thermocellum* xylanase 11A (PDB entry 1gmm) [19] is shown in purple.

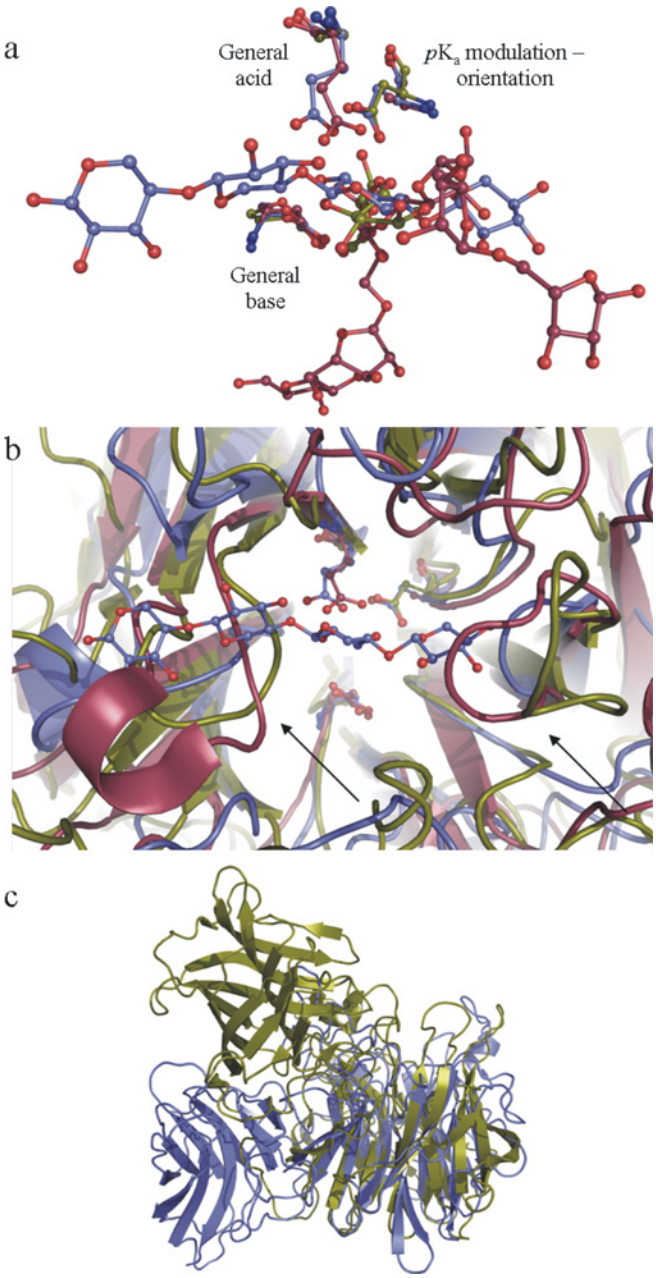

Figure 5 Structural differences at the active site of several GH family 43 members

(**a**) Detail of the superposition of the active site of *Bs*AXH-m2,3 (blue), α-L-arabinanase from *Ce. japonicus* (purple) and β-xylosidase from *G. stearothermophilus* (green) in complex with their substrate, showing the perpendicular orientation of the substrate. (**b**) Cartoon representation of the superposition of *Bs*AXH-m2,3 (blue) in complex with xylotetraose (ball and stick), α-L-arabinanase (purple) and β-xylosidase (green). The loop regions of α-L-arabinanase and β-xylosidase clashing with the xylotetraose are pointed. (**c**) Superposition of *Bs*AXH-m2,3 and β-xylosidase demonstrating the different positions of their CBM relative to the catalytic domain.

positioning [10], α-L-arabinanases and β-xylosidases display exo-activity, whereas AXH-m2,3 is a substituent-releasing enzyme.

Superposition of substrate-bound structures of α-L-arabinanase from *Ce. japonicus* (PDB entry 1gye) [15], β-xylosidase from *G. stearothermophilus* (PDB entry 2exk) [14] and *Bs*AXH-m2,3 in complex with xylotetraose reveals a different orientation of the sugar backbone. The xylan backbone bound to *Bs*AXH-m2,3 is perpendicular to the arabinose and xylose backbones bound to α-L-arabinanase and β-xylosidase respectively (Figure 5a). A closer look at the binding pocket of the three enzymes reveals a number of differences between them (Figure 5b). For α-L-arabinanase and β-xylosidase, the loop region connecting β-strands 2 and 3 of β-sheet II is much longer in comparison with the same loop region of *Bs*AXH-m2,3. In addition, their loop region connecting β-strands 2 and 3 of β-sheet IV is flipped in comparison with *Bs*AXH-m2,3. Thus, for α-L-arabinanase and β-xylosidase, the loop regions connecting β-strands 2 and 3 of

both β-sheets II and IV prevent the enzyme from hydrolysing substituents from the backbone.

Superposition of *Bs*AXH-m2,3 and β-xylosidase also reveals a different position of their CBMs with respect to the catalytic domain (Figure 5c). In β-xylosidase the CBM plays an important role in restricting the possible length of the incoming substrate by closing off a portion of the active site to form a pocket [14,17]. This results in an exo mode of action and the release of only one xylose unit, whereas for α-L-arabinanase, which has no CBM, arabinotriose units are released. For *Bs*AXH-m2,3, due to its position, the CBM does not seem to have an influence on the activity of the enzyme; there are no residues of its CBM that are involved in the active site or in the entrance to the active site. Therefore removing the CBM would probably not have an influence on the specificity of *Bs*AXH-m2,3.

In conclusion, soaking experiments of *Bs*AXH-m2,3 crystals with several sugars not only explained the specific activity of *Bs*AXH-m2,3, but also showed some remarkable differences in substrate binding between GH family 43 enzymes. Since substrate binding is mostly based on hydrophobic stacking interactions, this probably gives the necessary positional freedom to hydrolyse arabinose substituents either at the O-2 or O-3 hydroxyl position of the xylose. The observation of both exo-activity and the release of substituents within GH family 43 enzymes results from differences at the binding pocket which account for a different binding orientation of the sugar backbone.

ACKNOWLEDGEMENTS

We thank the staff of the EMBL/DESY Hamburg Outstation and the SLS for the provision of synchrotron facilities and skilful technical assistance.

FUNDING

This work was supported by the Flemish IWT (Instituut voor de aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen; IMPAXOS SBO project funding) [grant number 040073]. Financial support, through the I3 contract with the European Commission for support of access for external users of the EMBL/DESY Hamburg Outstation, was provided by the EMBL/DESY Hamburg Outstation and the SLS.

REFERENCES

1 Gray, K. A., Zhao, L. and Emptage, M. (2006) Bioethanol. Curr. Opin. Chem. Biol. **10**, 141–146
2 Saha, B. C. (2003) Hemicellulose bioconversion. J. Ind. Microbiol. Biotechnol. **30**, 279–291
3 Jeffries, T. W. (1994) Biodegradation of lignin and hemicelluloses. In Biochemistry of Microbial Degradation (Ratledge, C., ed.). pp. 233–277, Kluwer Academic Publishers, Dordrecht, The Netherlands
4 Pitson, S. M., Voragen, A. G. J. and Beldman, G. (1996) Stereochemical course of hydrolysis catalyzed by arabinofuranosyl hydrolases. FEBS Lett. **398**, 7–11
5 Kormelink, F. J. M., Searle-Van Leeuwen, M. J. F., Wood, T. M. and Voragen, A. G. J. (1991) (1,4)-beta-D-arabinoxylan arabinofuranohydrolase: a novel enzyme in the bioconversion of arabinoxylan. Appl. Microbiol. Biotechnol. **35**, 231–232
6 Laere, K. M. J. V., Beldman, G. and Voragen, A. G. J. (1997) A new arabinofuranohydrolase from *Bifidobacterium adolescentis* able to remove arabinosyl residues from double-substituted xylose units in arabinoxylan. Appl. Microbiol. Biotechnol. **47**, 231–235
7 Laere, K. M. J. V., Voragen, C. H. L., Kroef, T., Broek, L. A. M. V. d., Beldman, G. and Voragen, A. G. J. (1999) Purification and mode of action of two different arabinoxylan arabinofuranohydrolases from *Bifidobacterium adolescentis* DSM 20083. Appl. Microbiol. Biotechnol. **51**, 606–613
8 Ferré, H., Broberg, A., Duus, J. O. and Thomsen, K. K. (2000) A novel type of arabinoxylan arabinofuranohydrolase isolated from germinated barley—analysis of substrate preference and specificity by nano-probe NMR. Eur. J. Biochem. **267**, 6633–6641
9 Coutinho, P. M. and Henrissat, B. (1999) Carbohydrate-active enzymes: an integrated database approach. In Recent Advances in Carbohydrate Bioengineering (Gilbert, H. J., Davies, G., Henrissat, B. and Svensson, B., eds.), pp. 3–12, The Royal Society of Chemistry, Cambridge, U.K.
10 Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. Structure **3**, 853–859
11 Hövel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. and Schomburg, D. (2003) Crystal structure and snapshots along the reaction pathway of a family 51 alpha-L-arabinofuranosidase. EMBO J. **22**, 4922–4932
12 Miyanaga, A., Koseki, T., Matsuzawa, H., Wakagi, T., Shoun, H. and Fushinobu, S. (2004) Crystal structure of a family 54 alpha-L-arabinofuranosidase reveals a novel carbohydrate-binding module that can bind arabinose. J. Biol. Chem. **279**, 44907–44914
13 Bourgois, T. M., Van Craeyveld, V., Van Campenhout, S., Courtin, C. M., Delcour, J. A., Robben, J. and Volckaert, G. (2007) Recombinant expression and characterization of XynD from *Bacillus subtilis* subsp. *subtilis* ATCC 6051: a GH 43 arabinoxylan arabinofuranohydrolase. Appl. Microbiol. Biotechnol. **75**, 1309–1317
14 Brüx, C., Ben-David, A., Shallom-Shezifi, D., Leon, M., Karsten, N., Gil, S., Yuval, S. and Dietmar, S. (2006) The structure of an inverting GH43 beta-xylosidase from *Geobacillus stearothermophilus* with its substrate reveals the role of the three catalytic residues. J. Mol. Biol. **359**, 97–109
15 Nurizzo, D., Turkenburg, J. P., Charnock, S. J., Roberts, S. M., Dodson, E. J., McKie, V. A., Taylor, E. J., Gilbert, H. J. and Davies, G. J. (2002) *Cellvibrio japonicus* alpha-L-arabinase 43A has a novel five-blade beta-propeller fold. Nat. Struct. Biol. **9**, 665–668
16 Yamaguchi, A., Tada, T., Wada, K., Nakaniwa, T., Kitatani, T., Sogabe, Y., Takao, M., Sakai, T. and Nishimura, K. (2005) Structural basis for thermostability of endo-1,5-alpha-L-arabinanase from *Bacillus thermodenitrificans* TS-3. J. Biochem. **137**, 587–592
17 Brunzelle, J. S., Jordan, D. B., McCaslin, D. R., Olczak, A. and Wawrzak, Z. (2008) Structure of the two-subsite beta-D-xylosidase from *Selenomonas ruminantium* in complex with 1,3-bis[tris(hydroxymethyl)methylamino]propane. Arch. Biochem. Biophys. **474**, 157–166
18 Beisel, H.-G., Kawabata, S.-i., Iwanaga, S., Huber, R. and Bode, W. (1999) Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*. EMBO J. **18**, 2313–2322
19 Czjzek, M., Bolam, D. N., Mosbah, A., Allouch, J., Fontes, C. M. G. A., Ferreira, L. M. A., Bornet, O., Zamboni, V., Darbon, H., Smith, N. L. et al. (2001) The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. J. Biol. Chem. **276**, 48580–48587
20 Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. **382**, 769–781
21 Vandermarliere, E., Bourgois, T. M., Van Campenhout, S., Strelkov, S. V., Volckaert, G., Delcour, J. A., Courtin, C. M. and Rabijns, A. (2007) Crystallization and preliminary X-ray analysis of an arabinoxylan arabinofuranohydrolase from *Bacillus subtilis*. Acta Crystallogr. F **63**, 692–694
22 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology, Volume 276: Macromolecular Crystallography, Part A (Carter, C. W. and Sweet, J. R. M., eds.), pp. 307–326, Academic Press, New York
23 Leslie, A. (2006) The integration of macromolecular diffraction data. Acta Crystallogr. Sect. D Biol. Crystallogr. **62**, 48–57
24 Powell, H., Leslie, A. and Battye, G. (2007) Mosflm 7.0.1 and its new interface - iMosflm 0.5.3, In CCP4 Newsletter on Protein Crystallography, No. 46, Daresbury Laboratory, Warrington, U.K.
25 Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr. Sect. D Biol. Crystallogr. **62**, 72–82
26 Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. **50**, 760–763
27 Lammerts van Bueren, A., Morland, C., Gilbert, H. J. and Boraston, A. B. (2005) Family 6 carbohydrate binding modules recognize the non-reducing end of beta-1,3-linked glucans by presenting a unique ligand binding surface. J. Biol. Chem. **280**, 530–537
28 Keegan, R. M. and Winn, M. D. (2007) Automated search-model discovery and preparation for structure solution by molecular replacement. Acta Crystallogr. Sect. D Biol. Crystallogr. **63**, 447–457
29 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) *Phaser* crystallographic software. J. Appl. Crystallogr. **40**, 658–674
30 Perrakis, A., Morris, R. and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. **6**, 458–463
31 Jia-Xing, Y., Woolfson, M. M., Wilson, K. S. and Dodson, E. J. (2005) A modified ACORN to solve protein structures at resolutions of 1.7 Å or better. Acta Crystallogr. Sect. D Biol. Crystallogr. **61**, 1465–1475

32 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. **60**, 2126–2132
33 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of molecular structures by the maximum-likelihood method. Acta Crystallogr. Sect. D Biol. Crystallogr. **53**, 240–255
34 Lovell, S. C., Davis, I. W., Arendall, III, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. (2003) Structure validation by C-alpha geometry: phi, psi, and C-beta deviation. Protein Eng. Des. Sel. **50**, 437–450
35 Reference deleted
36 Davies, G. J., Wilson, K. S. and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. **321**, 557–559
37 Boraston, A. B., Notenboom, V., Warren, R. A. J., Kilburn, D. G., Rose, D. R. and Davies, G. J. (2003) Structure and ligand binding of carbohydrate-binding module *Cs*CBM6-3 reveals similarities with fucose-specific lectins and 'galactose-binding' domains. J. Mol. Biol. **327**, 659–669
38 Pires, V. M. R., Henshaw, J. L., Prates, J. A. M., Bolam, D. N., Ferreira, L. M. A., Fontes, C. M. G. A., Henrissat, B., Planas, A., Gilbert, H. J. and Czjzek, M. (2004) The crystal structure of the family 6 carbohydrate binding module from *Cellvibrio mixtus* endoglucanase 5A in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. J. Biol. Chem. **279**, 21560–21568
39 Henshaw, J. L., Bolam, D. N., Pires, V. M. R., Czjzek, M., Henrissat, B., Ferreira, L. M. A., Fontes, C. M. G. A. and Gilbert, H. J. (2004) The family 6 carbohydrate binding module *Cm*CBM6-2 contains two ligand-binding sites with distinct specificities. J. Biol. Chem. **279**, 21552–21559
40 Machovic, M. and Janecek, S. (2006) Starch-binding domains in the post-genomic era. Cell. Mol. Life Sci. **63**, 2710–2724
41 Feese, M. D., Kato, Y., Tamada, T., Kato, M., Komeda, T., Miura, Y., Hirose, M., Hondo, K., Kobayashi, K. and Kuroki, R. (2000) Crystal structure of glycosyltrehalose trehalohydrolase from the hyperthermophilic Archaeum *Sulfolobus solfataricus*. J. Mol. Biol. **301**, 451–464

Received 20 June 2008/27 October 2008; accepted 4 November 2008
Published as BJ Immediate Publication 4 November 2008, doi:10.1042/BJ20081256