Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structure–specificity relationships in Abp, a GH27 β-L-arabinopyranosidase from *Geobacillus stearothermophilus* T6

Shifra Lansky,[a] Rachel Salama,[b] Hodaya V. Solomon,[a] Hadar Feinberg,[a] Hassan Belrhali,[c] Yuval Shoham[b]* and Gil Shoham[a]*

[a]Institute of Chemistry and the Laboratory for Structural Chemistry and Biology, The Hebrew University of Jerusalem, Jerusalem 91904, Israel, [b]Department of Biotechnology and Food Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel, and [c]European Molecular Biology Laboratory, Grenoble Outstation and the Unit for Virus Host-Cell Interactions, Université Grenoble Alpes–EMBL–CNRS, 71 Avenue des Martyrs, 38000 Grenoble, France

Correspondence e-mail: yshoham@technion.ac.il, gil2@vms.huji.ac.il

Received 27 June 2014
Accepted 15 August 2014

PDB references: Abp, wild type, 4nx0; D197A mutant, 4nxk; D197A mutant, complex with ARB, 4nzf

L-Arabinose sugar residues are relatively abundant in plants and are found mainly in arabinan polysaccharides and in other arabinose-containing polysaccharides such as arabinoxylans and pectic arabinogalactans. The majority of the arabinose units in plants are present in the furanose form and only a small fraction of them are present in the pyranose form. The L-arabinan-utilization system in *Geobacillus stearothermophilus* T6, a Gram-positive thermophilic soil bacterium, has recently been characterized, and one of the key enzymes was found to be an intracellular β-L-arabinopyranosidase (Abp). Abp, a GH27 enzyme, was shown to remove β-L-arabinopyranose residues from synthetic substrates and from the native substrates sugar beet arabinan and larch arabinogalactan. The Abp monomer is made up of 448 amino acids, and based on sequence homology it was suggested that Asp197 is the catalytic nucleophile and Asp255 is the catalytic acid/base. In the current study, the detailed three-dimensional structure of wild-type Abp (at 2.28 Å resolution) and its catalytic mutant Abp-D197A with (at 2.20 Å resolution) and without (at 2.30 Å resolution) a bound L-arabinose product are reported as determined by X-ray crystallography. These structures demonstrate that the three-dimensional structure of the Abp monomer correlates with the general fold observed for GH27 proteins, consisting of two main domains: an N-terminal TIM-barrel domain and a C-terminal all-β domain. The two catalytic residues are located in the TIM-barrel domain, such that their carboxylic functional groups are about 5.9 Å from each other, consistent with a retaining mechanism. An isoleucine residue (Ile67) located at a key position in the active site is shown to play a critical role in the substrate specificity of Abp, providing a structural basis for the high preference of the enzyme towards arabinopyranoside over galactopyranoside substrates. The crystal structure demonstrates that Abp is a tetramer made up of two 'open-pincers' dimers, which clamp around each other to form a central cavity. The four active sites of the Abp tetramer are situated on the inner surface of this cavity, all opening into the central space of the cavity. The biological relevance of this tetrameric structure is supported by independent results obtained from size-exclusion chromatography (SEC), dynamic light-scattering (DLS) and small-angle X-ray scattering (SAXS) experiments. These data and their comparison to the structural data of related GH27 enzymes are used for a more general discussion concerning structure–selectivity aspects in this glycoside hydrolase (GH) family.

1. Introduction

Arabinan is a branched polysaccharide that forms part of pectin, which is the most complex polysaccharide in plant cell

walls, both structurally and functionally (Ermel *et al.*, 2000; Mohnen, 2008). The arabinan backbone consists of α-1,5-linked L-arabinofuranosyl (Araf) units and is decorated mainly with α-1,2- and α-1,3-linked arabinofuranosides (Shulami *et al.*, 2011). The majority of the arabinose units in plants are present in the furanose form (five-membered ring configuration); however, small amounts of the pyranose form (the six-membered ring configuration) are also present, mainly at the side-chain terminal ends of some polysaccharides. While the arabinopyranose form is favoured thermodynamically over the arabinofuranose form in solution (a ratio of about 90:10, respectively; Konishi *et al.*, 2007), the arabinopyranose moiety is relatively rare in natural polysaccharides. Reported values of the prevalence of arabinopyranose are up to 2% in arabinan (Shofiqur Rahman *et al.*, 2003; Cardoso *et al.*, 2002) and 5–8% in arabinogalactan (Willför *et al.*, 2002; Huisman *et al.*, 2001; Ponder & Richards, 1997).

β-L-Arabinopyranosidases (EC 3.2.1.88) remove a terminal β-L-arabinopyranose residue from the nonreducing end of arabino-oligosaccharides. The β-L-arabinopyranosidases studied and reported to date belong to glycoside hydrolase family 27 (GH27), which also includes α-galactosidases (EC 3.2.1.22), α-*N*-acetylgalactosaminidases (EC 3.2.1.49) and isomalto-dextranases (EC 3.2.1.94). All of the enzymes within this GH family act *via* a retaining mechanism (Davies & Henrissat, 1995), with two aspartic acids serving as the catalytic residues. To date, the crystal structures of nine GH27 enzymes have been determined, including α-galactosidases from human (*Homo sapiens*), rice (*Oryza sativa*), fungi (*Umbelopsis vinacea* and *Trichoderma reesei*) and yeast (*Saccharomyces cerevisiae*) (Kadaba *et al.*, 2008; Pinkett *et al.*, 2007; Hollenstein *et al.*, 2007; Gerber *et al.*, 2008; Ward *et al.*, 2007), α-*N*-acetylgalactosaminidases from human and chicken (*Gallus gallus*) (Clark & Garman, 2009; Garman *et al.*, 2002) and a β-L-arabinopyranosidase from a bacterium (*Streptomyces avermitilis*; Ichinose *et al.*, 2009). Of the 60 GH27 enzymes biochemically characterized to date, only four have been unequivocally confirmed as β-L-arabinopyranosidases. One of these enzymes is Abp, a GH27 β-L-arabinopyranosidase from *Geobacillus stearothermophilus* T6, which is the main focus of the structural study presented here.

G. stearothermophilus T6 is a Gram-positive thermophilic soil bacterium that possesses an extensive system for the utilization of plant cell-wall polysaccharides, including xylan, arabinan and galactan (Shulami *et al.*, 1999, 2007, 2011; Tabachnikov & Shoham, 2013). The bacterium produces a small number of endo-acting extracellular enzymes, which cleave the high-molecular-weight saccharide polymers in their immediate environment into short decorated oligosaccharides. These short oligosaccharides enter the bacterial cell *via* specialized ABC transporters (Böhm *et al.*, 2002; Rees *et al.*, 2009), and in the cell they are further hydrolyzed into the respective sugar monomers by an array of intracellular glycoside hydrolases (GHs). The most studied and best characterized of these complex utilization systems is that of xylan, in which the bacterium first secretes an extracellular xylanase (Teplitsky *et al.*, 1997, 2004; Bar *et al.*, 2004) that partially degrades external xylan into short decorated xylo-oligosaccharides. The degraded oligosaccharides are then transported into the cell *via* the corresponding ABC sugar transporters (Shulami *et al.*, 2007). Inside the cell, the decorated xylo-oligosaccharides are hydrolyzed by several dedicated side-chain-cleaving enzymes, including α-arabinofuranosidases (Shallom, Belakhov, Solomon, Gilead-Gropper *et al.*, 2002; Shallom, Belakhov, Solomon, Shoham *et al.*, 2002; Hövel, Shallom, Niefind, Baasov *et al.*, 2003; Hövel, Shallom, Niefind, Belakhov *et al.*, 2003), an α-glucuronidase (Teplitsky *et al.*, 1999; Zaide *et al.*, 2001; Golan *et al.*, 2004; Shallom *et al.*, 2004), acetyl-esterases (Alalouf *et al.*, 2011; Lansky, Alalouf *et al.*, 2013; Lansky, Alalouf, Salama *et al.*, 2014; Lansky, Alalouf, Solomon *et al.*, 2014) and finally by an intracellular xylanase (Teplitsky *et al.*, 2000; Solomon *et al.*, 2007) and several xylosidases (Bravman, Mechaly *et al.*, 2001; Bravman, Zolonitsky *et al.*, 2001, 2003; Bravman, Belakhov *et al.*, 2003; Shallom *et al.*, 2005; Brüx *et al.*, 2005, 2006; Ben-David *et al.*, 2007, 2008; Dann *et al.*, 2014).

The less studied L-arabinan-utilization system of *G. stearothermophilus* T6 has only recently been characterized in our laboratory (Shulami *et al.*, 2011). The system is located on a 38 kb gene segment and contains 23 genes including, amongst others, a gene for an oligo-arabinose-binding protein (AbnE) and genes for six glycoside hydrolases: extracellular and intracellular arabinanases (AbnA and AbnB, respectively; Alhassid *et al.*, 2009), two intracellular α-L-arabinofuranosidases (AbfA and AbfB; Gilead & Shoham, 1995; Shallom, Belakhov, Solomon, Gilead-Gropper *et al.*, 2002; Shallom, Belakhov, Solomon, Shoham *et al.*, 2002; Hövel, Shallom, Niefind, Belakhov *et al.*, 2003), an intracellular β-L-arabinofuranosidase (AraN; Lansky, Salama *et al.*, 2014) and the intracellular β-L-arabinopyranosidase (Abp) mentioned above (Fig. 1). The Abp enzyme has been biochemically characterized in our laboratory and shown to remove

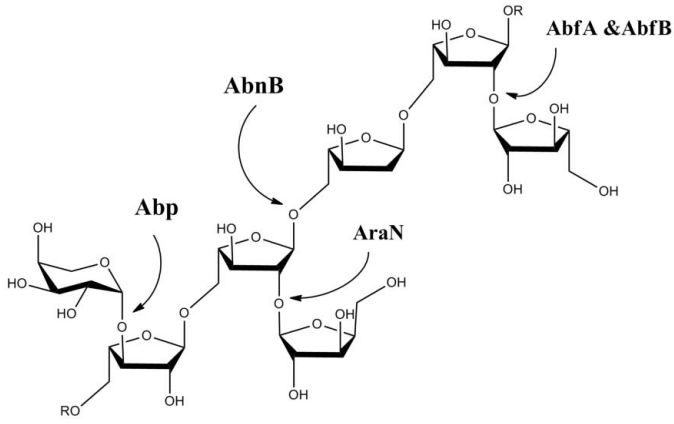


Figure 1
The common substrate for the intracellular glycoside hydrolases of the L-arabinan-utilization system of *G. stearothermophilus* is a short arabino-oligosaccharide containing α-L-arabinofuransoide moieties, AraN cleaves off the β-L-arabinofuranoside moieties and β-L-arabinopyranoside moieties. Abp cleaves off the β-L-arabinopyranoside moieties, AbfA and AbfB cleave off the α-L-arabinofuranoside moieties and AbnB cleaves the glycosidic bonds of the remaining linear oligosaccharide.

β-L-arabinopyranose residues from a synthetic substrate (pNP-β-L-arabinopyranoside) as well as from the native substrates sugar beet arabinan and larch arabinogalactan (Salama *et al.*, 2012).

The Abp monomer is made up of 448 amino acids, with a calculated molecular mass of 51 178 Da. Based on sequence homologies with other GH27 enzymes, it has been suggested that Asp197 is the catalytic nucleophile and Asp255 is the catalytic acid/base (Salama *et al.*, 2012). Recently, we were able to crystallize the wild-type enzyme (Abp-WT) and several of its catalytic mutants, as well as collect full X-ray diffraction data from some of these crystals (Lansky, Salama *et al.*, 2013). In the present paper, we report the structural information resulting from these diffraction data, specifically the crystal structures of Abp-WT and the Abp-D197A catalytic mutant, and also the complex of Abp-D197A with L-arabinose, one of its reaction products. These structures have been determined at 2.28, 2.30 and 2.20 Å resolution, respectively. The crystal structure of Abp is the second reported structure of an enzyme with confirmed β-L-arabinopyranosidase catalytic activity. The detailed three-dimensional structure presented here, together with its structural comparisons with selected members of GH27, should therefore enable a better understanding of the catalytic mechanisms and substrate-binding modes of β-L-arabinopyranosidases and related GH27 enzymes.

2. Experimental methods

2.1. Purification and oligomeric characterization of Abp-WT

Expression and purification of wild-type Abp (Abp-WT; including the His tag) were carried out as described previously (Lansky, Salama *et al.*, 2013). Briefly, the N-terminally His-tagged protein was isolated on a HisTrap column, resulting in a distinct peak, which was collected and dialyzed against 50 m*M* Tris–HCl buffer pH 7.0, 100 m*M* NaCl, 0.02% sodium azide. The resulting protein solution was analyzed by size-exclusion chromatography (SEC) and dynamic light scattering (DLS) in order to evaluate the oligomeric state of the native protein.

2.1.1. Size-exclusion chromatography. The apparent molecular weight of Abp-WT in solution was estimated by SEC using an ÄKTA explorer system (Pharmacia) equipped with a Superose 12 HR gel-filtration column (GE Healthcare Life Sciences) of 24 ml total column volume. Protein samples (100 µl) were applied onto the column and eluted at room temperature with a solution consisting of 50 m*M* Tris–HCl buffer pH 7.0, 100 m*M* NaCl, 0.02% sodium azide at a flow rate of 0.5 ml min^{-1}. Molecular weights were determined from regression analysis of the log relative molecular weight (M_r) of protein standards as a function of the available partition coefficient (K_{av}). The void volume, 6.85 ml, was determined using dextran blue. The protein standards used (all from *G. stearothermophilus*) included the intracellular xylanase XT6 (43 800 Da), the extracellular β-1,4-galactanase GanA (87 000 Da), the intracellular β-galactosidase GanB (240 000 Da) and the intracellular xylosidase XynB2 (160 000 Da). Using this experimental setup, the Abp-WT protein eluted from the size-exclusion column as a single peak with an estimated molecular mass of about 210 kDa (data not shown), indicating that under these conditions Abp-WT is most likely a tetramer in solution.

Table 1
Representative crystallographic data-collection parameters for Abp.

Values in parentheses are for the outer resolution shell.

Protein	Abp-WT (medium resolution)	Abp-WT (high resolution)	Abp-D197A	Abp-D197A–ARB
Wavelength (Å)	0.954	0.954	0.954	0.954
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters				
a (Å)	107.5	107.7	107.7	107.6
b (Å)	201.9	202.2	203.5	201.5
c (Å)	286.4	287.3	287.0	286.9
Resolution range (Å)	50.00–2.90 (2.95–2.90)	30.00–2.28 (2.32–2.28)	25.00–2.30 (2.34–2.30)	35.00–2.20 (2.24–2.20)
Multiplicity	5.8 (4.1)	5.9 (3.4)	5.5 (3.7)	3.5 (2.4)
$\langle I/\sigma(I)\rangle$	7.0 (2.8)	9.6 (4.4)	8.3 (4.0)	11.5 (2.7)
Mosaicity (°)	0.366	0.181	0.274	0.354
Completeness (%)	98.8 (95.0)	99.8 (97.6)	97.1 (79.1)	89.6 (83.4)
R_{merge}† (%)	11.9 (37.3)	7.3 (25.2)	8.6 (32.5)	7.3 (25.0)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of observation i of reflection hkl.

2.1.2. Dynamic light scattering (DLS). DLS analysis was performed using a Malvern Nano ZetaSizer model ZEN 3600 (Malvern Instruments Ltd, UK) with a purified protein concentration of 5 mg ml^{-1} (0.097 m*M*) at 25°C. Analysis of the results was performed using the *ZetaSizer* software v.7.03 provided by Malvern Instruments. Based on this analysis, a single oligomeric state of Abp-WT was detected in solution (Supplementary Fig. S1[1]), with an estimated hydrodynamic diameter of 9.3 nm (an average of 7.9 nm calculated by number and of 10.7 nm calculated by volume). These results are consistent with the tetrameric form of Abp-WT determined in the current crystal structure, as discussed below.

2.2. Crystal structure determination of Abp-WT

Two data sets at medium and high resolution were collected from single crystals of Abp-WT as described in detail elsewhere (Lansky, Salama *et al.*, 2013; see Supporting Information). Representative statistics and processing details are summarized in Table 1. The initial three-dimensional structure of Abp-WT was determined by molecular-replacement (MR) techniques using *Phaser* (McCoy *et al.*, 2007) as implemented in the *PHENIX* software suite (Adams *et al.*, 2010). The structure of a highly homologous protein (65% amino-acid sequence identity), the GH27 enzyme BH1870 from *Bacillus halodurans* (PDB entry 3cc1; Joint Center for Structural Genomics, unpublished work), was used as a search model.

At the time of the initial structure-determination process, only the medium-resolution X-ray data (2.90 Å resolution)

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: DZ5341).

Table 2
Representative structure-determination and refinement parameters for Abp.

Protein	Abp-WT (medium resolution)	Abp-WT (high resolution)	Abp-D197A	Abp-D197A–ARB
Model refinement				
Data resolution range (Å)	30.00–2.90	30.00–2.28	24.86–2.30	35.00–2.20
R factor† (%)	19.5	14.6	16.9	17.7
No. of reflections	138737	261171	255755	268240
R_{free}‡ (%)	23.2	17.7	20.0	21.3
No. of reflections	6992	14365	14105	16025
Refined model				
No. of residues	3437	3446	3447	3450
No. of atoms	28455	32729	31524	31510
No. of water molecules	569	4204	3062	2932
No. of molecules in asymmetric unit	8	8	8	8
Average *B* factor (Å^2)				
Protein	34.7	26.9	34.6	35.1
Solvent	16.0	36.8	37.3	37.3
Ligands	51.5	63.1	59.3	56.3
R.m.s. deviation				
Bond lengths (Å)	0.020	0.018	0.019	0.017
Bond angles (°)	2.10	1.89	1.84	1.84
Luzzati estimated coordinate error (Å)	0.36	0.21	0.25	0.25
Ramachandran plot, residues in (%)				
Favoured region	94.3	96.8	97.2	97.3
Allowed region	4.2	3.2	2.7	2.7
Outlier region	1.5	0	0.1	0
PDB code	—	4nx0	4nxk	4nzf

† R factor = $\sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$. ‡ A random subset (5.0%) of the data was used for the calculation of R_{free}.

were available for Abp-WT (Lansky, Salama *et al.*, 2013), and it was these data that were used for the first structure-determination cycles. *Phaser* (McCoy *et al.*, 2007) gave a relevant MR solution with eight independent Abp monomers (labelled *A*–*H*) in the crystallographic asymmetric unit, in agreement with the asymmetric unit content estimated by the reported V_M calculations (Lansky, Salama *et al.*, 2013). The coordinates and maps obtained from this solution were then used as input files for the *AutoBuild Wizard* program (Terwilliger *et al.*, 2008), as available within the *PHENIX* suite (Adams *et al.*, 2010). The resulting eight-monomer model of the asymmetric unit demonstrated a good fit to the calculated electron-density maps, with reasonable intermediate values of the *R* factor (29%) and R_{free} (33%).

However, the maps still had regions with smeared and non-interpretable electron density, especially around residues 72–84, 104–107, 231–254, 273–282, 364–372, 400–405 and 421–444, indicating that these regions needed to be rebuilt. This procedure was performed manually with *Coot* (Emsley *et al.*, 2010) together with concomitant refinement with *REFMAC*5 (Murshudov *et al.*, 2011; using NCS restraints), and greatly improved the quality of the electron-density maps. This allowed a slow reconstruction of almost all amino-acid residues of the eight protein monomers in the asymmetric unit. In the last stages of refinement, *ARP*/*wARP* (Langer *et al.*, 2008) was used to fit water molecules to peaks in the ($F_o - F_c$) difference electron-density maps, based on suitable distances and orientations for forming hydrogen bonds with the corresponding protein residues. A similar procedure was used to fit various other protein-bound molecules (such as glycerol molecules and sulfate ions) originating from the crystallization and crystal-cooling solutions. Refinement of this model resulted in a final *R* factor of 19.5% and R_{free} of 23.2%. Although it had reasonable crystallographic parameters (Table 2), the structure obtained at 2.9 Å resolution had numerous sections of unclear conformations, leading us to collect an additional synchrotron data set, this time at an improved resolution of 2.28 Å (BM14 beamline, ESRF, Grenoble, France; Lansky, Salama *et al.*, 2013).

MR calculations with *MOLREP* (Vagin & Teplyakov, 2010) gave a clear and unequivocal solution for the higher resolution data, with initial *R* factor and R_{free} values of 22.1 and 25.5%, respectively, using the 2.9 Å resolution Abp-WT structure as a starting model. The greatly improved electron-density maps were then used for final corrections and refinement of the Abp-WT model, in a similar way to the procedure described above. The refinement converged to a final *R* factor of 14.6% and a final R_{free} of 17.7%. Representative parameters of the refinement and the final Abp-WT model at 2.28 Å resolution are listed in Table 2.

2.3. Content and quality of the final crystallographic model of Abp-WT

The asymmetric unit of the current structure contains eight independent molecules of the Abp-WT monomer, labelled here as chains *A*, *B*, *C*, *D*, *E*, *F*, *G* and *H*. The first 14–15 residues of the N-termini of all eight chains and the last four residues of the C-termini of most of the chains (except for chain *D*) were not included in the final model, as the corresponding experimental electron density was too weak to define their positions with high confidence. Two *cis*-peptide bonds were identified between residues Ser231 and Pro232 and between residues Gly233 and Pro234. A total of 4204 water molecules were identified and refined in the crystallographic asymmetric unit (including eight monomers of the protein) of the final Abp-WT structure. Additional significant electron densities in the crystallographic ($F_o - F_c$) difference maps were assigned as five citrate anions, 40 glycerol molecules and 85 sulfate anions (per asymmetric unit). The total content of the final crystallographic asymmetric unit of the Abp-WT structure, as refined in the present work, includes eight independent protein monomers, with 27 795 (non-H) protein atoms and 4934 (non-H) nonprotein atoms, resulting in a total of 32 729 non-H atoms per asymmetric unit (Table 2).

PROCHECK (Laskowski *et al.*, 1993) was used for validation of all structural parameters and stereochemical

calculations (Table 2). The Ramachandran plot (Ramachandran *et al.*, 1963) for Abp-WT showed that 96.8% of the residues were in most favoured regions, 3.2% were in additionally allowed regions and no residues were in disallowed regions. The Abp protein molecules conformed closely to standard bond lengths and bond angles, as defined by Engh & Huber (1991), with estimated root-mean-square (r.m.s.) deviations of about 0.018 Å and 1.89°, respectively. The overall average *B* factor for the protein atoms was 26.9 Å^2, and based on the resolution of the current data (2.28 Å) and the final *R* factor obtained (14.6%), the average experimental error in the coordinates of the final model was about 0.21 Å as estimated by the Luzzati error analysis (Luzzati, 1952). Such a coordinate-error range is quite reasonable for protein structures of this resolution and asymmetric unit content, permitting a meaningful and reliable analysis of the interactions and geometries involved in the detailed three-dimensional structure presented here.

2.4. The Abp-D197A catalytic mutant with and without arabinose

Expression, purification, crystallization and preliminary crystallographic analysis of the nucleophile catalytic mutant of Abp (Abp-D197A) were performed in a similar manner to the Abp-WT protein, as described previously (Lansky, Salama *et al.*, 2013). The Abp-D197A crystals were found to be closely isomorphous to those of Abp-WT in both space group ($P2_12_12_1$) and unit-cell parameters (Table 1). The fully refined three-dimensional structure of Abp-WT at 2.28 Å resolution (see above) was used as the basis for structure determination

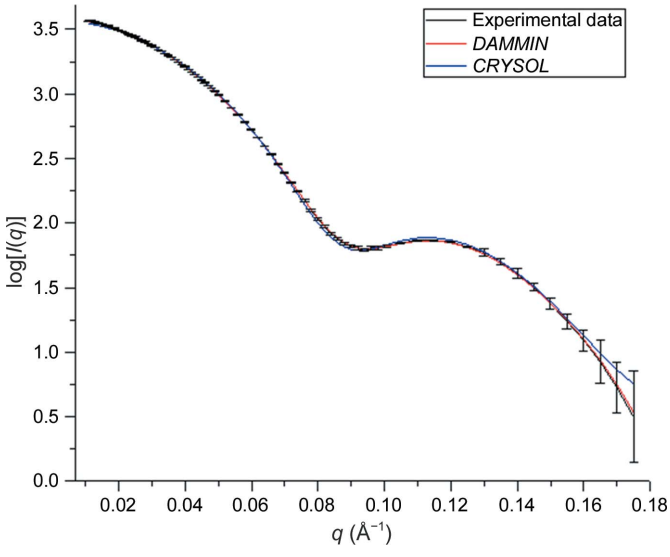


Figure 2
SAXS experimental results for the Abp-WT protein in solution. The scattering curve for Abp-WT: log[$I(q)$] as a function of the momentum-transfer vector q. The q range used for data analysis was q = 0.011–0.175 Å^{-1}. Experimental data are shown in black. The red curve corresponds to the simulated scattering curve from a molecular-envelope model of *DAMMIN*, χ^2 = 1.254. The blue curve corresponds to the simulated scattering curve from the crystallographic tetramer model of Abp presented in this paper, calculated by *CRYSOL*, χ = 2.843.

of the Abp-D197A mutant at 2.30 Å resolution. Model building and refinement were performed as described above for the Abp-WT structure, resulting in a final *R* factor of 16.9% and a final $R_{\rm free}$ of 20.0%. The final geometrical parameters of the Abp-D197A structure are very similar to those of the Abp-WT structure, including the estimated average coordinate error (Table 2).

The complex of Abp-D197A with β-L-arabinose (Abp-D197A–ARB) was obtained in the crystal by a short soaking procedure. A fully grown Abp-D197A crystal was soaked in a solution consisting of 4 m*M* β-L-arabinose, 18% glycerol, 1.35 *M* ammonium sulfate, 0.09 *M* sodium citrate buffer pH 5 for about 60 s prior to flash-cooling. Crystallographic diffraction data for the complex were collected from this crystal to a 2.20 Å resolution limit, using the same general procedure and the same crystallographic setup as described previously for Abp-WT (Lansky, Salama *et al.*, 2013). The structure of the Abp-D197A–ARB complex was determined using the refined Abp-D197A structure described above. Model building and refinement were performed as described above for the Abp-WT structure, resulting in a final *R* factor of 17.7%, a final $R_{\rm free}$ of 21.3% and reasonable geometrical statistics similar to those of the structures of Abp-WT and Abp-D197A. Representative data-collection, structure-determination and refinement parameters for the Abp-D197A–ARB complex are summarized in Tables 1 and 2.

2.5. SAXS experiments with Abp-WT

2.5.1. SAXS data collection. Small-angle X-ray scattering (SAXS) data were measured for Abp-WT on the X9 beamline at the National Synchrotron Light Source (NSLS; Brookhaven National Laboratory, Upton, New York, USA). Protein samples were prepared in a buffer consisting of 50 m*M* Tris–HCl pH 7.0, 100 m*M* NaCl, 0.02% sodium azide at a protein concentration of ∼10 mg ml^{-1}. For buffer-scattering subtraction, identical buffer samples were prepared and their scattering was measured immediately before measurement of the corresponding protein samples. Each measurement consisted of a 30 s exposure to a 400 × 200 µm X-ray beam (wavelength of 0.918 Å) of a 20 µl sample flowing continuously during the X-ray exposure through a 1 mm diameter capillary. The measurements were repeated three times for accuracy. The measured scattered intensity ranged from q = 0.008 Å^{-1} to q = 1.810 Å^{-1}. The small-angle scattering intensity, $I(q)$, was measured using a PILATUS 300K detector located 3.4 m from the sample for the small-angle part of the scattering. A Photonic Science CCD detector positioned 0.47 m from the sample (Yang, 2013) was used for the wide-angle part of the scattering (Allaire & Yang, 2011). The overlapping region (q = 0.12–1.8 Å^{-1}) was used to merge these two data sets using the in-house Python script-based software developed at NSLS X9 (*pyXS*; Allaire & Yang, 2011).

2.5.2. SAXS data analysis. Data analysis was carried out using the in-house *pyXS* software (Allaire & Yang, 2011) and the SAXS software package *ATSAS* (Konarev *et al.*, 2006; Petoukhov *et al.*, 2012). The net protein-scattering data of the

sample, $I(q)$, were generated by subtraction of the scattering of the same buffer solution from the scattering pattern of the protein solution using *pyXS*. Owing to a partial buffer mismatch between the protein sample and the buffer solution at wider angles, it was decided not to include these data in the subsequent analysis, and only the data from low-angle scattering (q = 0.011–0.175 Å^{-1}) were used for the final scattering curve (Fig. 2). The Guinier plot for Abp-WT (qR_g = 0.58–1.28) showed relatively good linearity (adjusted R^2 = 0.9995; Supplementary Fig. S2*a*), indicating no aggregation of the protein sample. Using the Guinier plot approximation [$I(q) = I_0\exp(-q^2R_g^2/3)$], where a plot of $I(q)$ and q^2 is linear for $q < 1.3/R_g$ (Guinier, 1939), the radius of gyration (R_g) of the protein was calculated to be 41.4 ± 0.4 Å and I_0 to be 3938 ± 6. *PRIMUS* (Konarev *et al.*, 2003) was used for data truncation and generation of the pair-distance distribution function $P(r)$, as shown in Supplementary Fig. S2(*b*).

30 independent *ab initio* models for the molecular envelope of Abp were constructed from the $P(r)$ function using *DAMMIN* (Svergun, 1999). The resulting models, which were generated with no imposition of structure or symmetry, show a good fit to the experimental data, with χ^2 values ranging between 1.25 and 1.38 for each of the 30 models. Improvement of these initial models was performed with *DAMAVER* (Volkov & Svergun, 2003), in which the 30 *ab initio* models were averaged and filtered to yield the final molecular-envelope model for Abp-WT presented below. Superimposition of the SAXS-based models with the crystallographic atomic structure of Abp-WT was performed with *SUPCOMB* (Kozin & Svergun, 2001), and fitting of the experimental SAXS curves to the theoretical curves calculated from the crystal structures was performed using *CRYSOL* (Svergun *et al.*, 1995).

2.6. Preparation and characterization of the Abp-I67D mutant

As previously suggested (Salama *et al.*, 2012) and further discussed below, Ile67 appeared to be a key residue for the substrate-binding and catalytic specificity of Abp. In order to test this hypothesis, this residue was replaced by aspartic acid, as briefly summarized in the following. Site-directed mutagenesis of the *abp* gene was performed using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, California, USA). The mutagenic primers for the I67D replacement were as follows (the mutated nucleotides are shown in bold): 5′-GAATATGTTGTGGTCGAT**GAT**CAGTGGTATGAACCCGGC-3′ and 5′-GCCGGGTTCATACCACTG**ATC**ATCGACCACAACATATTC-3′. The replacement was confirmed by DNA sequencing. The mutated protein was overexpressed and purified as for the wild type. The activity towards synthetic aryl-glycoside substrates was measured by following the release of *p*-nitrophenol at 405 nm (extinction coefficient = 12 125 M^{-1} cm^{-1}). Kinetic studies were performed at 40°C in 100 m*M* citric acid, 200 m*M* Na_2HPO_4 buffer pH 6.0 containing 1 mg ml^{-1} BSA with the synthetic substrates *p*NP-β-L-arabinopyranoside (*p*NP-β-L-AraP) and *p*NP-α-D-galactopyranoside (*p*NP-α-D-GalP) at concentration ranges of 0.2–4.6 and 0.3–60 m*M*, respectively. The reactions were initiated by adding 20 µl of appropriately diluted prewarmed enzyme solution to 180 µl prewarmed reaction buffer and were terminated by adding 50 µl 0.2 *M* Na_2CO_3. For the substrate *p*NP-α-D-GalP V_{max} was not achievable and the catalytic efficiency, k_{cat}/K_m, was hence estimated only at low substrate concentrations ([S] << K_m).

2.7. Calculations and figure preparation

The matrices for the superposition of the different structures were calculated with the *PDBeFold* web server (Krissinel & Henrick, 2004). Calculations of surface areas and solvation free-energy gain values (Δ^iG) were performed with the *PISA* web server (Krissinel & Henrick, 2007). Calculations of electrostatic surface potential according to the linearized Poisson–Boltzmann equation were performed with the *APBS* plug-in v.1.3 (Baker *et al.*, 2001) as implemented in *PyMOL* (v.1.5.0.4; Schrödinger). Figs. 3, 4(*b*), 4(*c*), 5, 7(*b*), 7(*d*), 8, 9, 10(*a*) and 10(*c*) and Supplementary Figs. S3, S4 and S5 were prepared using *Chimera* (Pettersen *et al.*, 2004); Figs. 4(*a*), 6 and 10(*b*) were prepared using *PyMOL*; Fig. 2 and Supplementary Fig. S2 were prepared using *Origin* (OriginLab, Northampton, Massachusetts, USA); and Figs. 1, 7(*a*) and 7(*c*) were prepared using *ChemDraw* 3*D*.

2.8. PDB accession codes

The atomic coordinates of Abp-WT, Abp-D197A and Abp-D197A–ARB have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (Berman *et al.*, 2000) under accession codes 4nx0, 4nxk and 4nzf, respectively.

3. Results and discussion

3.1. The overall structure of Abp

3.1.1. The Abp monomer. Of the structures presented here, the structure of Abp-WT is the most relevant for functional conclusions and it will therefore be used as the main reference for the detailed structural analysis presented below. The Abp protein belongs to glycoside hydrolase family 27 (GH27), and accordingly its tertiary structure is similar to the general fold observed previously for the available protein structures that belong to this family [for example, PDB entries 3cc1 (Joint Center for Structural Genomics, unpublished work), 1uas (Fujimoto *et al.*, 2003), 3a5v (Fujimoto *et al.*, 2009), 1ktb (Garman *et al.*, 2002), 3h53 (Clark & Garman, 2009), 1r46 (Garman & Garboczi, 2004), 1szn (Golubev *et al.*, 2004) and 3lrk (Fernández-Leiro *et al.*, 2010)]. This fold usually consists of two main domains: an N-terminal TIM-barrel domain and a C-terminal all-β domain (Fig. 3).

The TIM-barrel domain of Abp (residues 1–357) is similar to the general $(\beta/\alpha)_8$ fold which is observed in many glycoside hydrolases. It is made up of eight parallel β-sheets arranged in a circle in the middle of the protein, surrounded by eight long α-helices in the outer wall of the barrel, with two extra short

α-helices and seven 3_{10}-helices in between (Fig. 3 and Supplementary Fig. S3). The eight β-sheets are formed by residues 27–30 (β1), 62–65 (β2), 122–128 (β3), 193–197 (β4), 227–230 (β5), 249–251 (β6), 282–284 (β7) and 328–330 (β8). The eight long α-helices are formed by residues 40–55 (α2), 109–119 (α3), 174–190 (α5), 208–222 (α6), 239–247 (α7), 259–277 (α8), 310–325 (α9) and 338–357 (α10). The two short α-helices are formed by residues 31–36 (α1) and 133–139 (α4). The seven 3_{10}-helices are formed by residues 16–23 (h1), 56–59 (h2), 100–104 (h3), 148–152 (h4), 236–238 (h5), 298–302 (h6) and 333–337 (h7).

The C-terminal domain (residues 358–448) is made of eight antiparallel β-strands containing two 'Greek-key' motifs. The strands are formed by residues 358–366 (β9), 369–375 (β10), 381–387 (β11), 393–398 (β12), 408–412 (β13), 417–422 (β14), 425–430 (β15) and 435–442 (β16). Although the exact borders between the two domains are not very clear, they seem to be connected by a short polypeptide 'hinge' made of residues Gly357, Tyr358 and Gly359 (Fig. 3 and Supplementary Fig. S3). The buried surface between the TIM-barrel and the all-β C-terminal domains is 1264 Å^2, corresponding to a solvation free-energy gain (Δ^iG) of -13.9 kcal mol^{-1} and a Δ^iG *P*-value of 0.115 (as calculated by the *PISA* server; Krissinel & Henrick, 2007).

The TIM-barrel domain is the domain that is most probably responsible for the actual enzymatic catalysis, as the active site of Abp is situated in a pocket in the middle of this domain (Fig. 3). This is well correlated with sequence alignment with related members of the GH27 family and is confirmed by the location of the bound L-arabinose which is the product of the catalytic reaction (see below). In contrast, the function of the all-β C-terminal domain is as yet unknown. In a previous study on human α-galactosidase (PDB entry 3hg4; Guce *et al.*, 2010), a related GH27 enzyme, an α-galactose molecule was found bound to a second ligand-binding site at the interface between these two domains. However, in Abp and other homologous GH27 structures no ligands were found bound to this area, or to anywhere else in the all-β domain, questioning the possibility that this domain is functionally involved in additional or alternative substrate binding.

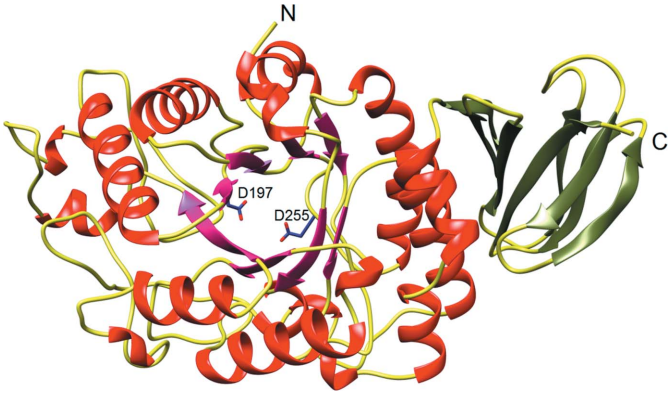


Figure 3
The overall structure of the Abp monomer, showing the catalytic N-terminal TIM-barrel domain (red) and the C-terminal all-β domain (green). The catalytic aspartic residues (Asp197 and Asp255) are shown in blue.

In order to examine the specific role of the all-β C-terminal domain in Abp, an attempt was made to remove this domain and check the effect of its removal on the structure and function of the resulting truncated protein (Abp-Δ358–448; see Supporting Information). Nevertheless, when trying to express the Abp protein without this domain, the resulting protein became insoluble and aggregated within inclusion bodies, and it was not possible to purify it in a soluble (and active) form. A possible reason for this behaviour of the Abp-Δ358–448 deletion mutant could, at least in part, be related to the exposure of hydrophobic residues at the original interface between the two domains of Abp-WT. Such hydrophobic amino-acid side chains (*e.g.* Trp260, Leu263, Phe267, Leu320, Ile323 and Val353), which are involved in hydrophobic/aromatic interactions at the inter-domain contact area in the Abp-WT protein, become exposed to solvent in the deletion mutant, leading to immediate protein aggregation and precipitation. In principle, these surface residues could be replaced by charged or polar residues, thereby providing a more adequate aqueous interface and potentially preventing protein precipitation (without direct effects on the active site and catalysis). Such experiments are currently under way in our laboratory.

3.1.2. The Abp tetramer. As stated, the asymmetric unit of the current crystal structure of Abp includes eight independent monomers labelled *A*–*H*. These eight monomers are arranged as two tetrameric assemblies lying side by side, where one tetramer is built of monomers *A*, *B*, *C* and *D* and the other tetramer is built of monomers *E*, *F*, *G* and *H*. Superposition of these two independent tetramers reveals that they are practically identical, with an r.m.s.d. value of 0.67 Å calculated over the 1718 C$^\alpha$ atoms involved (as calculated by the *PDBeFold* server; Krissinel & Henrick, 2004). The tetramer composed of chains *ABCD* will be used in the analysis and discussion below.

The relevance of such a tetrameric arrangement is supported by information obtained from size-exclusion chromatography (data not shown) and dynamic light scattering (Supplementary Fig. S1), clearly indicating that Abp is a tetramer in solution. Moreover, in both analyses a single, relatively sharp, peak was obtained for a 'physiological' solution of the Abp protein, confirming that the Abp tetramer is the only oligomeric form of the protein in such a solution, at least at the detection limits of these two techniques. This tetrameric quaternary structure of Abp is further supported by the protein-assembly calculations performed using the *PISA* web server (Krissinel & Henrick, 2007) and by small-angle X-ray scattering (SAXS) experiments. The SAXS data produced a molecular envelope, independent of the crystallographic structure, which showed a relatively good fit when superimposed upon the tetramer composed of chains *ABCD* (Fig. 4*a*; Supporting Information). Furthermore, upon comparing the experimental scattering curve of Abp {log[$I(q)$] *versus* q} with the simulated scattering curve obtained from the crystallographic tetramer of chains *ABCD* (calculated with

CRYSOL; Svergun *et al.*, 1995), a very good fit was achieved (Fig. 2).

The Abp tetramer consists of a 'dimer of dimers', in which one dimer is formed by chains *A* and *C* and the other by chains

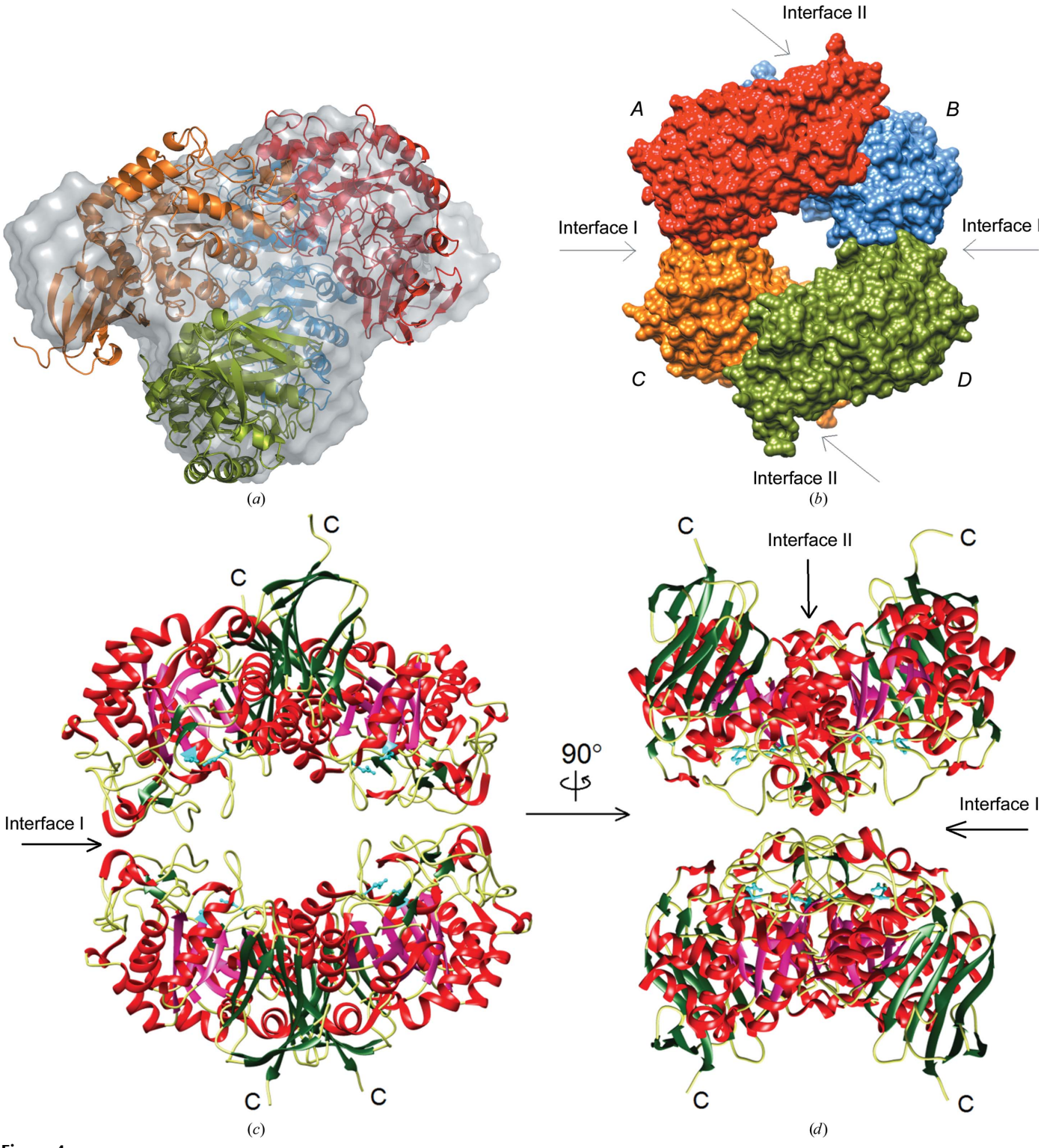


Figure 4
The Abp tetramer. (*a*) Superposition of the molecular envelope of Abp calculated from the SAXS data on the crystallographic model of the Abp tetramer. The superposition demonstrates a very good fit between the SAXS and the crystallographic results, confirming the tetrameric assembly of Abp in solution. (*b*) Surface representation of the Abp tetramer (chains *A*–*D*). Chain *A* is in red, chain *B* in blue, chain *C* in orange and chain *D* in green. The tetramer is built of two dimers (chains *A* and *C* and chains *B* and *D*) 'clamped' around each other. (*c*, *d*) Ribbon representation of two different orientations of the Abp tetramer (in *d* the view is 90° to the view of *c*). The secondary structure of the TIM-barrel domain is shown in red and pink, and the secondary structure of the all-β domain is shown in green. The catalytic residues are shown in turquoise.

B and *D*. These dimers may be pictured as two 'open pincers', which clamp around each other to form a central cavity in the middle (Figs. 4*b* and 4*c*). The four active sites of Abp are situated at the inner surface of this cavity (Fig. 4*c*), all opening into the central space. The two active sites within each dimer are separated by ∼30 Å, while these two active-site pairs are separated by ∼40 Å. The resulting homotetramer possesses D_2 symmetry, containing three twofold symmetry axes that intersect at the centre. The overall surface area of the Abp tetramer is about 59 560 Å^2 and the total buried area is about 15 540 Å^2. The smallest cross-section of the crystallographic tetramer is ∼65 Å, as measured between Ser393*A* and Ser393*B*, the largest cross-section of the crystallographic tetramer is ∼100 Å, as measured between Ala443*B* and Ala443*C*, and the third cross-section is ∼86 Å as measured between Ser102*A* and Ser102*B*, resulting in an average tetrameric cross-section of ∼84 Å. These values correlate reasonably well with the corresponding values obtained by the DLS analysis for the Abp protein in solution (see above).

There are two types of intermolecular interfaces in the tetramer (see Supporting Information). Interface type I is the contact interface between the monomers that form each of the dimers (*i.e.* between chains *A* and *C* and similarly between chains *B* and *D*; Figs. 4*b*, 4*c* and 5*a*). Interface type II is the contact interface formed by the interaction of the two dimers with one another (*i.e.* between chains *A* and *B* and between chains *C* and *D*; Figs. 4*b* and 5*b*). The protein–protein contacts involved in interface I (within the dimer *A* and *C* and within the dimer *B* and *D*) include four hydrogen bonds and two π-stacking interactions, all between relatively conserved residues within the GH27 family. Specifically, for the contact surface involved in the *A* and *C* dimer, the hydrogen bonds are between Arg133*A* NH1 and Tyr166*C* OH and between Arg150*A* NH2 and Phe82*C* O, and similarly for their corresponding mates between Arg133*C* NH1 and Tyr166*A* OH and between Arg150*C* NH2 and Phe82*A* O. The π-stacking interactions are formed between Tyr166*A* and Phe82*C* and similarly for their corresponding mates between Tyr166*C* and Phe82*A* (Fig. 5*a*).

The protein–protein contacts involved in interface II (between chains *A* and *B* and between chains *C* and *D*) involve six hydrogen bonds and four salt-bridge interactions, all between less-conserved residues in the GH27 family. Specifically, for the *A* and *B* interface the hydrogen bonds are between Thr241*A* OG1 and Tyr271*B* O, between Thr241*A* OG1 and Lys272*B* O and between Lys272*A* NZ and Asp238*B* O, and similarly for their corresponding mates between Thr241*B* OG1 and Tyr271*A* O, between Thr241*B* OG1 and Lys272*A* O and between Lys272*B* NZ and Asp238*A* O. The salt-bridge interactions are formed between Glu245*A* OE2 and Lys275*B* NZ and between Lys272*A* NZ and Asp238*B* OD1, and similarly between Glu245*B* OE2 and Lys275*A* NZ and between Lys272*B* NZ and Asp238*A* OD1 (Fig. 5*b*). All of these interactions are summarized in Table 3 together with the corresponding distances.

The reasons for the particular oligomeric arrangement of the Abp protein are as yet unclear, and could be related to various parameters such as overall stability, selective substrate accessibility and/or preferred active-site positioning (Lansky, Alalouf, Soloman *et al.*, 2014). Another possible reason for the specific oligomerization observed may be related, directly or indirectly, to electrostatic surface considerations. Looking at the electrostatic surface potential of the Abp tetramer (Fig. 6*a*), one can see a relative concentration of negative potential lining the inner cavity of the protein, where the active sites are situated. This observation is in accordance with the negative potential 'patches' that are usually associated with active sites

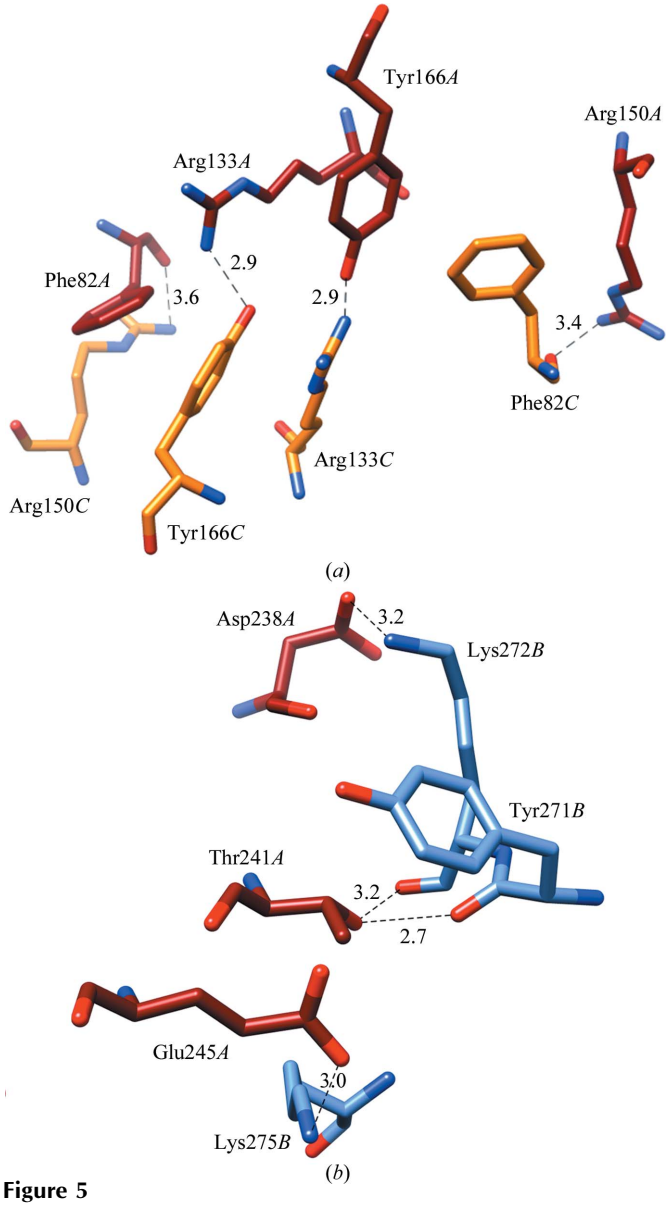


Figure 5
Intermolecular contacts within the Abp tetramer. (*a*) A close-up of the main molecular contacts involved in interface type I (contacts forming the basic dimers), specifically the interactions between chains *A* (red) and *C* (orange). These contacts are relatively conserved among GH27 proteins. (*b*) A close-up of the main molecular contacts involved in interface type II (contacts between the basic dimers), specifically between chains *A* (red) and *B* (blue). These contacts are relatively nonconserved within GH27 proteins. Only half of the interactions are presented, as the other contacts at each interface are between the corresponding mates of the amino acids shown. The distances are given in Å.

Table 3
Summary of the intermolecular contacts involved in the Abp tetramer.

Type of contact	Residues in contact	Atoms in contact	Distance (Å)
Contacts forming interface I			
Hydrogen bond	Arg133*A*–Tyr166*C*	NH1–OH	2.87
	Arg133*B*–Tyr166*D*	NH1–OH	2.89
	Arg133*C*–Tyr166*A*	NH1–OH	2.88
	Arg133*D*–Tyr166*B*	NH1–OH	2.90
Hydrogen bond	Arg150*A*–Phe82*C*	NH2–O	3.64
	Arg150*B*–Phe82*D*	NH2–O	3.57
	Arg150*C*–Phe82*A*	NH2–O	3.41
	Arg150*B*–Phe82*D*	NH2–O	3.54
π-stacking	Phe82*A*–Tyr166*C*	—	—
	Phe82*B*–Tyr166*D*	—	—
	Phe82*C*–Tyr166*A*	—	—
	Phe82*D*–Tyr166*B*	—	—
Contacts forming interface II			
Hydrogen bond	Thr241*A*–Tyr271*B*	OG1–O	2.70
	Thr241*B*–Tyr271*A*	OG1–O	2.72
	Thr241*C*–Tyr271*D*	OG1–O	2.62
	Thr241*D*–Tyr271*C*	OG1–O	2.70
Hydrogen bond	Thr241*A*–Lys272*B*	OG1–O	3.22
	Thr241*B*–Lys272*A*	OG1–O	3.25
	Thr241*C*–Lys272*D*	OG1–O	3.32
	Thr241*D*–Lys272*C*	OG1–O	3.30
Salt bridge	Glu245*A*–Lys275*B*	OE2–NZ	2.97
	Glu245*B*–Lys275*A*	OE2–NZ	3.10
	Glu245*C*–Lys275*D*	OE2–NZ	3.05
	Glu245*D*–Lys275*C*	OE2–NZ	3.07
Salt bridge	Lys272*A*–Asp238*B*	NZ–OD1	3.13
Hydrogen bond		NZ–O	3.09
Salt bridge	Lys272*B*–Asp238*A*	NZ–OD1	3.12
Hydrogen bond		NZ–O	3.20
Salt bridge	Lys272*C*–Asp238*D*	NZ–OD1	2.97
Hydrogen bond		NZ–O	3.06
Salt bridge	Lys272*D*–Asp238*C*	NZ–OD1	3.06
Hydrogen bond		NZ–O	3.01

of glycoside hydrolases (Stawiski *et al.*, 2002) and have been claimed to provide specific advantages in attracting sugar-like substrates to the active-site area. In the current case of Abp, the special tetrameric assembly seems to amplify this negative potential by concentrating it all into one area (Fig. 6*a*), as the negative potential 'patches' calculated for an Abp monomer or an Abp dimer are much less pronounced (Fig. 6*b*). Such oligomerically amplified negative potential 'patches' may have functional consequences (Stawiski *et al.*, 2002), as they can possibly help attract the substrate towards the right area of the enzyme and thus improve the catalytic activity of the enzyme.

3.2. The structures of Abp-D197A and Abp-D197A–ARB

The crystal structures of Abp-D197A and Abp-D197A–ARB are very similar to that of Abp-WT, as described above. Similarly to Abp-WT, the eight independent monomers of both structures are practically identical to each other, as reflected in r.m.s.d. values of 0.09–0.34 Å between the eight monomers of the Abp-D197A structure and r.m.s.d. values of 0.19–0.38 Å between the eight monomers of the Abp-D197A–ARB structure. The refined structures of Abp-D197A and Abp-D197A–ARB do not deviate significantly from the structure of the native protein (Abp-WT), except for the obvious differences in the mutation site (in both structures) and the substrate/product-binding site (in the complex). The r.m.s.d. values between the monomers of Abp-WT and

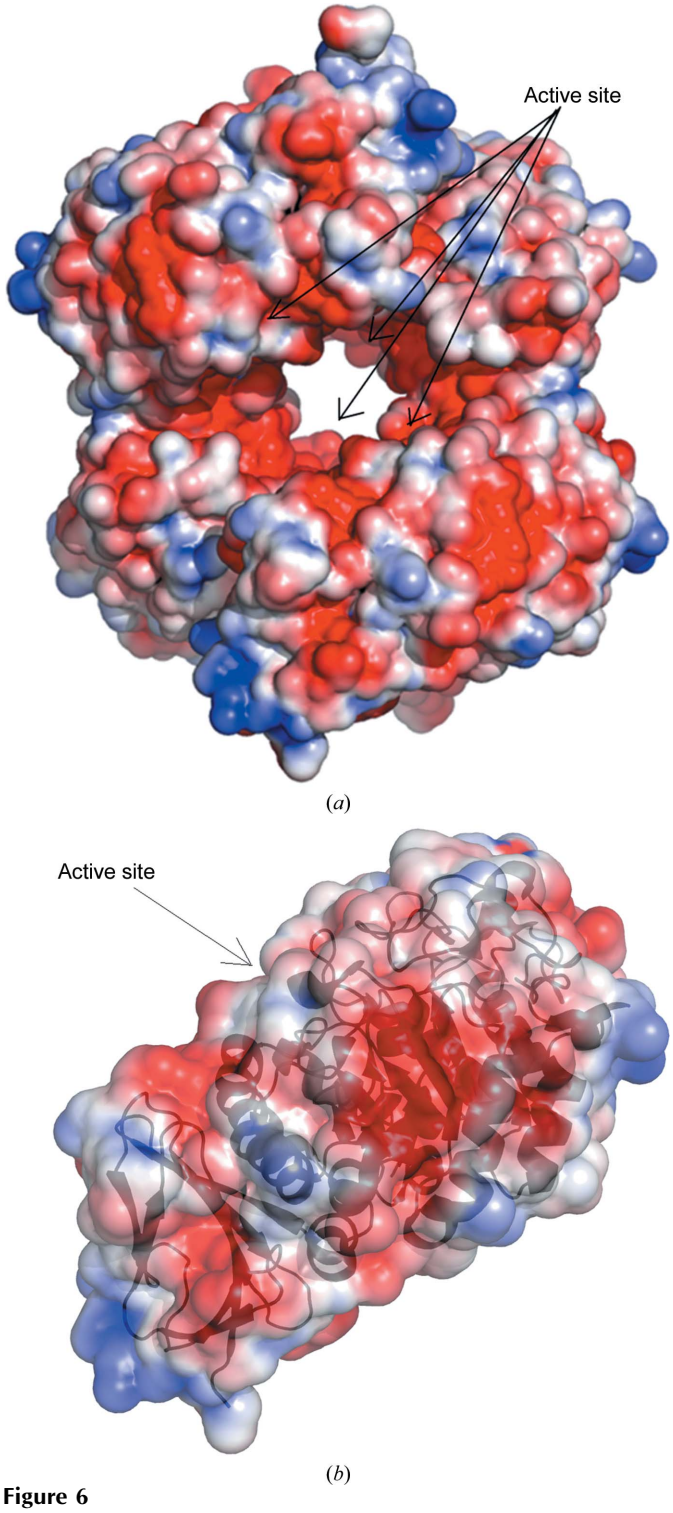


Figure 6
Electrostatic surface potential of Abp-WT (calculated with the *APBS* plug-in as implemented in *PyMOL*). The potential is in accordance with the linearized Poisson–Boltzmann equation, using the PARSE force field, pH 7 and a solvent dielectric parameter of 78.0. The potential gradient is in the range $+3kT$/e to $-3kT$/e, where dark blue represents the most positive potential and dark red represents the most negative potential. Lower potential levels are shown in lighter shades of the corresponding colour. (*a*) The potential calculated for the Abp tetramer, demonstrating a 'patch' of negative potential at the inner surface of the tetramer around the general location of the active sites. (*b*) The electrostatic potential calculated in the same manner for the Abp monomer, demonstrating a significantly reduced negative potential around the active site.

Abp-D197A are in the range 0.09–0.32 Å and the r.m.s.d. values between the monomers of Abp-WT and Abp-D197A–ARB are in the range 0.08–0.30 Å, demonstrating very similar tertiary structures. Comparing the crystallographic tetramers in the three crystal structures, the resulting r.m.s.d. values are 0.14 Å between Abp-WT and Abp-D197A and 0.13 Å between Abp-WT and Abp-D197A–ARB, confirming that they are also very similar in their quaternary structures (Supplementary Fig. S4). Moreover, even the number and the positions of solvent molecules modeled in the three crystal structures are quite similar, making the additional β-L-arabinose molecules in the structure of Abp-D197A–ARB the main difference between the three structural models (see below). As such, these monomers and tetramers similarities between the different structures further confirm the validity, conformational stability and biological significance of the Abp crystal structures presented here.

3.3. The active site

As mentioned above, Abp is a β-L-arabinopyranosidase that removes β-L-arabinopyranoside moieties from the side chains of L-arabinose-containing polysaccharides (Fig. 1; Salama *et al.*, 2012). Sequence-alignment analysis determined that the catalytic residues of Abp are Asp197 and Asp255, with Asp197 serving as the nucleophile and Asp255 as the acid/base for the reaction (Salama *et al.*, 2012). The distance between these two residues in the structure of Abp-WT is about 5.9 Å, confirming that Abp operates *via* the retaining mechanism of glycoside hydrolases (Davies & Henrissat, 1995; Fig. 7*a*), as typical distances between the catalytic residues for this mechanism are usually in the range of 5–6 Å (McCarter & Withers, 1994).

As described above, β-L-arabinopyranoside molecules were captured in the active sites of Abp-D197A–ARB (Fig. 7*b*), a state which corresponds to step VI of the proposed catalytic cycle of Abp (Fig. 7*a*). The Abp-D197A–ARB structure thus provides additional information regarding this step. The arabinopyranoside monomer appeared to be bound in subsite −1 (Davies *et al.*, 1997), which is the site that binds the sugar ring situated next to the cleavage point on the nonreducing side. As seen for other GH structures, this is usually the site that presents the strongest interactions with bound sugar rings, either alone or in the framework of a longer sugar oligomer. In the structures of Abp-WT and Abp-D197A, a glycerol molecule was found to be bound in this site, practically mimicking the position and interactions of 'half' of the sugar molecule found in the same place in the Abp-D197A–ARB complex (Supplementary Fig. S5). Interestingly, in all three structures presented here an additional glycerol molecule was found bound to a second binding site in the active site (Fig. 7*b*). As for the first glycerol molecule, this glycerol probably mimics a bound sugar molecule in subsite +1, *i.e.* the sugar monomer on the reducing side of the cleaved bond. In the case of a true substrate of Abp, this site is likely to bind the adjacent sugar monomer composing the arabinose-containing polysaccharide.

Table 4
Residues involved in the binding sites of Abp.

The interactions are deduced from the Abp-D197A–ARB complex. As a result of the mutation, all the interactions involving Asp197 are missing.

Residue and atom	Type of interaction	Atoms of ligand in contact	Averaged distance (Å)
Subsite −1 (bound arabinose)			
Ser231 OG	Hydrogen bond	O1	3.2
Pro232 N	Hydrogen bond	O1	3.4
Arg251 NH1, NH2	Hydrogen bond	O2	2.9, 3.3
		O3	3.4
Asp255 OD1	Hydrogen bond	O2	2.6
Lys195 NZ	Hydrogen bond	O3	3.0
		O4	2.9
Asp286 OD2	Water-mediated hydrogen bond	O3	2.7, 2.8
Asp66 OD1, OD2	Hydrogen bond	O4	2.5, 3.3
His126 NE2	Hydrogen bond	O4	3.1
		O5	3.6
Ile67	Hydrophobic interaction	C4, C5	—
Trp32	Hydrophobic interaction	C3, C4, C5, O3	—
Subsite +1 (bound glycerol)			
Asp254 OD1	Hydrogen bond	O2	2.8
Asp255 OD1, OD2	Hydrogen bond	O1	2.7, 3.4
Trp161	Hydrophobic interaction	C1, C2, C3	—

According to the structure of the Abp-D197A–ARB complex, the residues involved in binding the β-L-arabinose molecule in subsite −1 of Abp include Trp32, Asp66, Ile67, His126, Lys195, Asp197, Ser231, Pro232, Arg251 and Asp255. Specifically, Ser231 OG and Pro232 N form hydrogen bonds to O1 of the bound arabinose, Asp255 OD1 and Arg251 NH1 and NH2 form hydrogen bonds to O2 of the arabinose, Arg251 NH1 and Lys195 NZ form hydrogen bonds to O3, Asp66 OD1 and OD2, Lys195 NZ and His126 NE2 form hydrogen bonds to O4, and His126 NE2 forms a hydrogen bond to O5 of the bound arabinose (Figs. 7*b* and 7*c*). In addition, Ile67 and Trp32 seem to be involved in hydrophobic interactions with the flat face of the arabinose ring made up of its C3, C4, C5 and O3 atoms, and Asp286 forms a water-mediated hydrogen bond to O3 of the bound arabinose. The residues involved in binding the glycerol molecule in the putative subsite +1 are Trp161, Asp254 and Asp255. Specifically, Asp254 and Asp255 form hydrogen bonds to two of the glycerol O atoms, and Trp161 is involved in hydrophobic interactions with the C atoms of the bound glycerol (Figs. 7*b* and 7*c*). Except for Asp254, which is only semi-conserved, all of these residues are highly conserved within the GH27 family, as evaluated by the *ConSurf* server (Ashkenazy *et al.*, 2010; Supplementary Fig. S6). Such high-level conservation indicates that the observed interactions are biologically significant, especially those in the −1 subsite. The distances of these important contacts (averaged over all eight binding sites) are summarized in Table 4.

3.4. The catalytic mode of action of Abp

Interestingly, the eight arabinose molecules found in the eight independent active sites of the crystallographic asymmetric unit have been captured in slightly different confor-

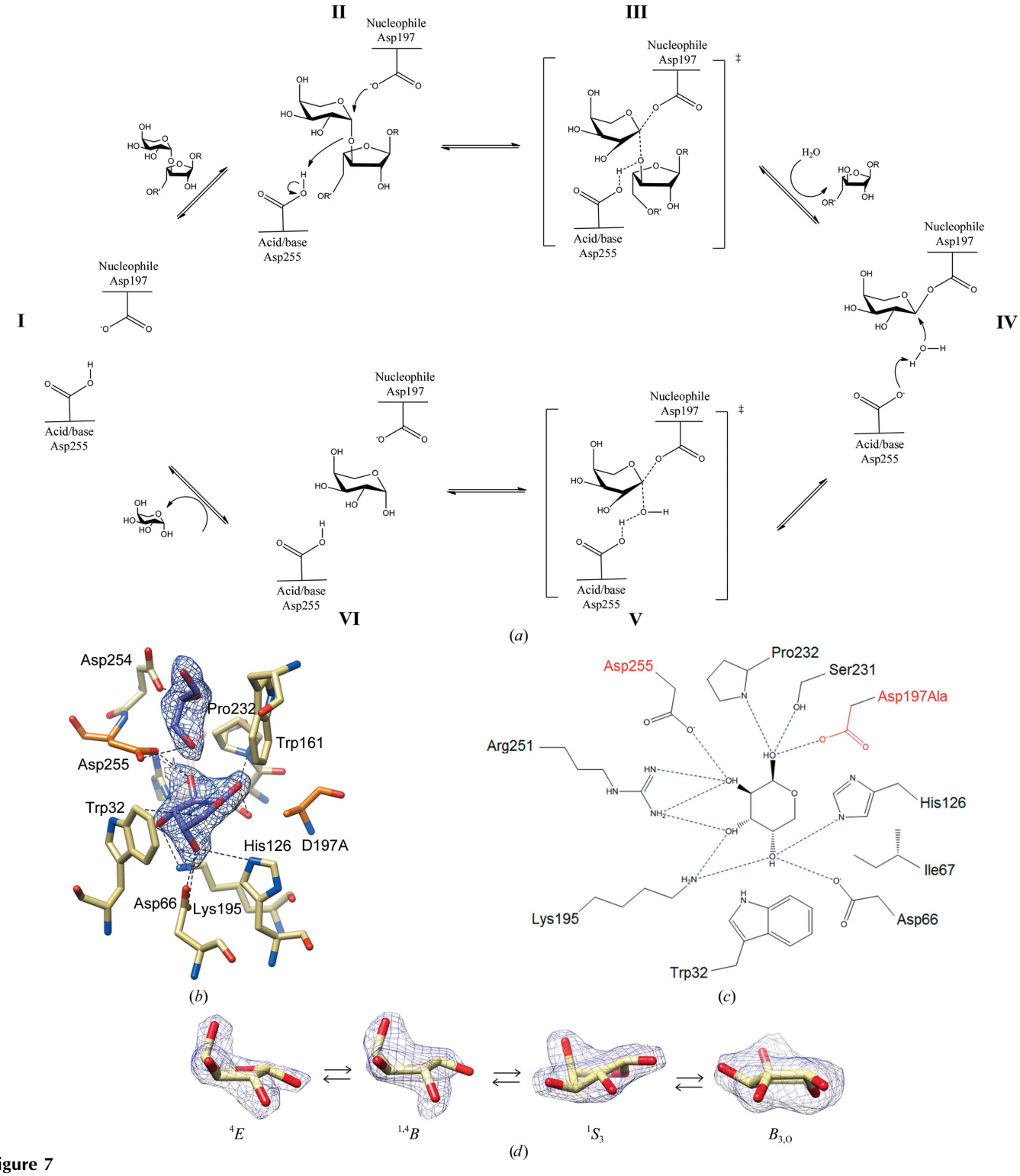


Figure 7
The active site and catalytic mechanism of Abp. (*a*) Proposed catalytic cycle of Abp, corresponding to the retaining mechanism of glycoside hydrolases. (I) Active site without substrate. (II) Nucleophilic attack of Asp197 on the anomeric centre of β-L-arabinopyranoside, catalyzed by the acidic Asp255. (III) Tetrahedral transition state, resulting in release of the arabino-oligosaccharide from the β-L-arabinopyranoside moiety. (IV) Intermediate state of Abp, in which the β-L-arabinopyranoside moiety is covalently bound to Asp197. In this step a water molecule attacks the anomeric centre of the bound sugar, catalyzed by Asp255, which acts here as a base. (V) Tetrahedral transition state. (VI) The bound β-L-arabinopyranoside is released into the active site and exits. (*b*) The active site of the Abp-D197A–ARB complex, showing the β-L-arabinose molecule captured in the −1 subsite and the glycerol molecule captured in the +1 subsite. The experimental electron density around the arabinose and glycerol molecules corresponds to a $(2F_o - F_c)$ map at 2.2 Å resolution contoured at the 1.5σ level. The catalytic residues are shown in orange; the nucleophile Asp197 was mutated to Ala. (*c*) A scheme of the active site of Abp-D197A–ARB, showing the residues forming interactions with the bound β-L-arabinose molecule. The carboxylate group of the native Asp197 is modelled in red. (*d*) The different conformations of the β-L-arabinose molecules found trapped in the eight crystallographically independent active sites of Abp. Each of these molecules is shown together with the corresponding $(2F_o - F_c)$ electron-density section contoured at the 1.3σ level.

mations. These conformations include an envelope conformation ^{4}E (in chains *A* and *E*), a skew-boat conformation $^{1}S_{3}$ (in chains *B*, *C*, *D* and *G*) and two boat conformations, $^{1,4}B$ (chain *F*) and $B_{3,O}$ (chain *H*) (IUPAC–IUB Joint Commission on Biochemical Nomenclature, 1980; Fig. 7*d*). According to the Cremer–Pople sphere (Cremer & Pople, 1975), these conformations convert directly into each other, so that ^{4}E converts into $^{1,4}B$, which converts into $^{1}S_{3}$ on its way to $B_{3,O}$ (Davies *et al.*, 2012; Fig. 7*d*). Hence, these conformations seem to hint at the different conformational changes that the β-L-arabinose molecule undergoes while being cleaved from the polysaccharide substrate by Abp (Davies *et al.*, 2012), which probably allows better nucleophilic access to the anomeric carbon reaction centre of the sugar (Knott *et al.*, 2014). Indeed, the $^{1}S_{3}$ skew-boat conformation matches the covalent intermediate trapped in the crystal structure of the GH27 human α-galactosidase (PDB entry 3hg4; Guce *et al.*, 2010). The $^{1}S_{3}$ conformation also corresponds to the itinerary which was found in other enzymes of the GH27 family (Davies *et al.*, 2012), while the other conformations found in Abp seem to hint at additional conformations in this pathway. In this respect, it should be noted that the conformational differences indicated above, although definitely significant, are quite close to the experimental resolution limits. Hence, for a more conclusive interpretation these initial observations should be further investigated with a series of substrates in different experimental conditions and, if possible, should be followed at significantly higher resolutions.

Another interesting aspect emanating from the present structures of Abp involves the particular substrate specificity

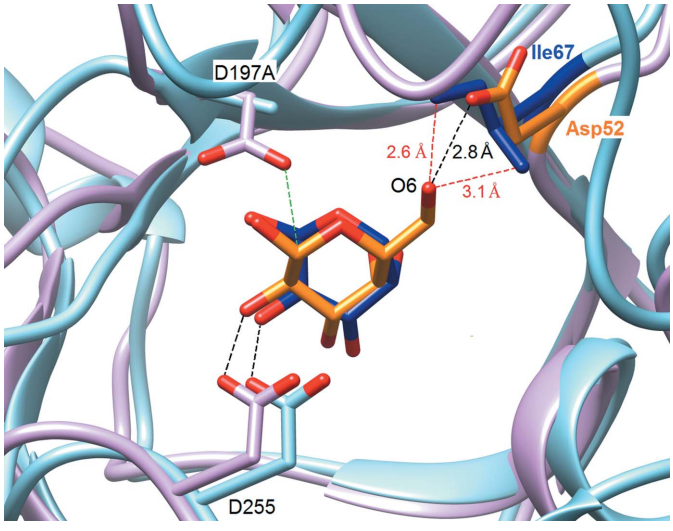


Figure 8
A superposition of the active site of Abp-D197A–ARB (containing a bound arabinose molecule) with the active site of rice-α-Gal, which contains a bound galactose molecule (PDB entry 1uas). Abp is shown in light blue and rice-α-Gal is shown in pink. The arabinose and galactose molecules are shown in dark blue and orange, respectively. It is demonstrated that in rice-α-Gal Asp52 (orange) forms a hydrogen bond to O6 of the bound galactose, while in Abp this residue is replaced by Ile67 (blue), which cannot form a similar interaction. This difference may govern the specificity of Abp towards pNP-β-L-arabinopyranoside compared with pNP-α-D-galactopyranoside substrates.

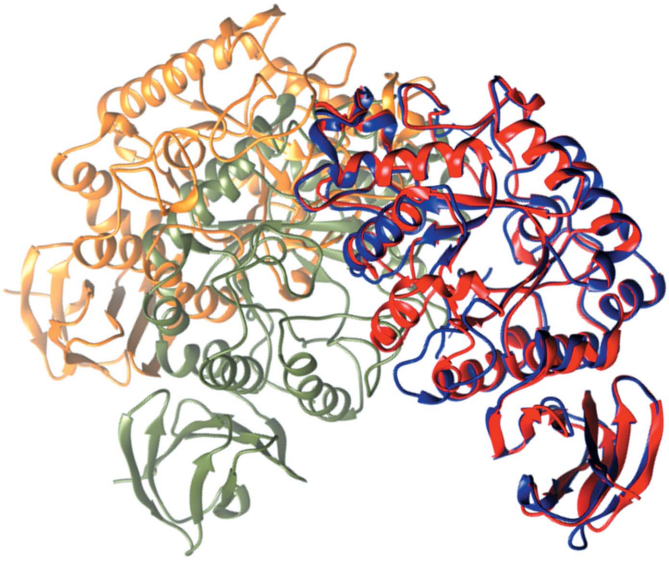
Table 5
The catalytic activity of Abp-WT and Abp-I67D.

*p*NP-β-L-AraP, pNP-β-L-arabinopyranoside; *p*NP-α-D-GalP, pNP-α-D-galactopyranoside; nd, not determined.

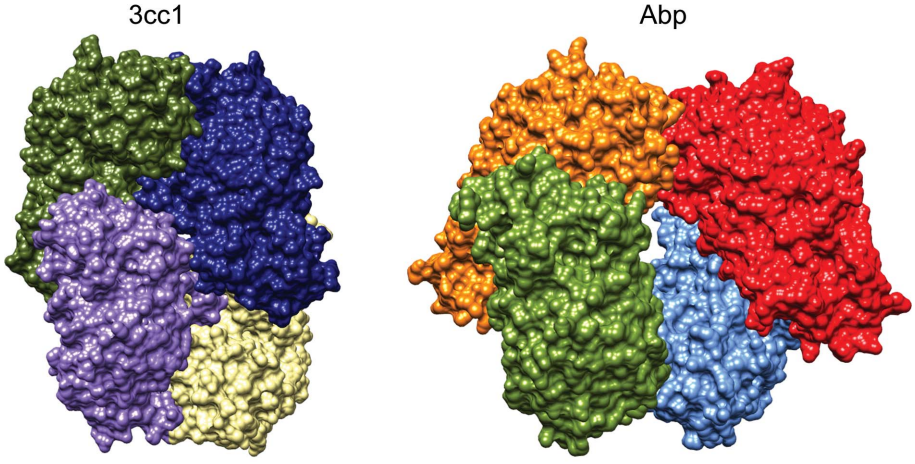
	*p*NP-β-L-AraP			*p*NP-α-D-GalP		
	k_{cat} (s^{-1})	K_m (m*M*)	k_{cat}/K_m (s^{-1} mM^{-1})	k_{cat} (s^{-1})	K_m (m*M*)	k_{cat}/K_m (s^{-1} mM^{-1})
Abp-WT	3.78	0.75	5.04	nd	nd	$\sim 2 \times 10^{-5}$
Abp-I67D	1.58	0.85	1.86	nd	nd	$\sim 1 \times 10^{-2}$

of the enzyme. Previous studies have shown that Abp is about four orders of magnitude more active towards pNP-β-L-arabinofuranoside than towards pNP-α-D-galactopyranoside (Salama *et al.*, 2012). Such highly selective substrate specificity may be explained, at least in part, by comparing the binding mode of β-L-arabinose in the Abp-D197A–ARB structure with the corresponding binding mode of D-galactose in the structure of rice-α-galactosidase (rice-α-Gal; PDB entry 1uas; Fujimoto *et al.*, 2003). In the rice-α-Gal structure, O6 of the bound galactose forms a tight hydrogen bond to residue Asp52 in the active site of the enzyme. This Asp residue is highly conserved among galactosidases of the GH27 family, but is surprisingly absent in Abp, where the analogous residue is Ile67. Obviously, the Ile67 residue in Abp cannot form this important hydrogen bond owing to the chemical nature of the side chain. Moreover, the exact position of Ile67 in the active site of Abp indicates that the side chain of this residue would be involved in a hydrophobic clash with a hypothetically bound galactose in this part of the active site (Fig. 8). In addition, in Abp the actual space available at the −1 subsite is not suitable for the larger galactose molecule in comparison with the slightly smaller arabinose molecule, especially with respect to the extra C and O atoms (C6 and O6) of the galactose molecule, making the binding of the smaller arabinose molecule preferable. These specific structural parameters seem to account for the observed preference of Abp towards arabino substrates compared with galacto substrates.

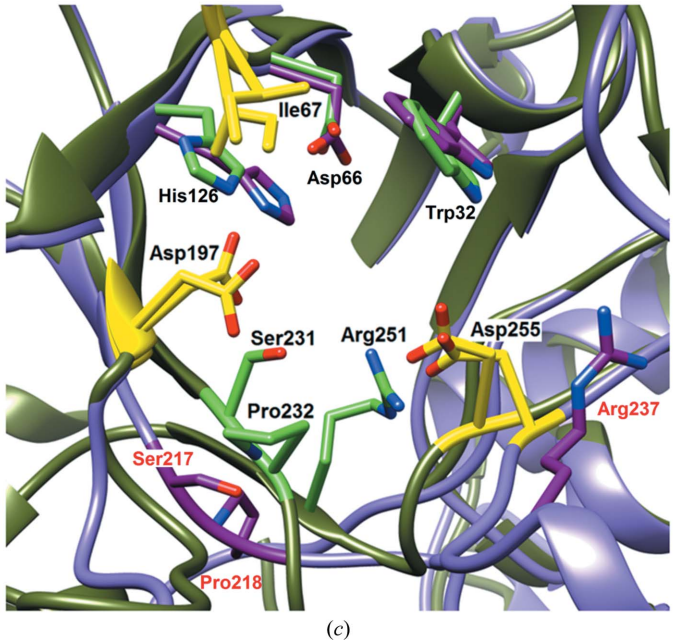
In order to test the proposed effect of residue 67 in Abp, we prepared a specific mutant of the enzyme in which the active-site residue Ile67 was changed to Asp (Abp-I67D). This mutant indeed showed an increased activity towards pNP-α-D-galactopyranoside (*p*NP-α-D-GalP), while only a small decrease was observed in its activity towards pNP-β-L-arabinopyranoside (*p*NP-β-L-AraP). Specifically, the kinetic parameters towards *p*NP-β-L-AraP were derived from typical Michaelis–Menten curves (data not shown), resulting in k_{cat} values of 3.78 and 1.58 s^{-1} and K_m values of 0.75 and 0.85 m*M* for Abp-WT and Abp-I67D, respectively (Table 5). These results indicate that the site-specific I67D replacement did not dramatically affect the activity of Abp towards *p*NP-β-L-AraP, suggesting that the differences in shape and charge involved in the Ile→Asp mutation do not interfere significantly with the key interactions of the WT enzyme with arabino substrates. The catalytic constants towards *p*NP-α-D-GalP could not be obtained independently (the high K_m was beyond the measurement limits); however, the specificity constant,



(*a*)



(*b*)



(*c*)

k_{cat}/K_m, could be estimated at low substrate concentrations, resulting in values of $\sim 2 \times 10^{-5}$ and $\sim 1 \times 10^{-2}$ s^{-1} mM^{-1} for Abp-WT and Abp-I67D, respectively (Table 5). Thus, a single replacement of an isoleucine residue by an aspartic acid at position 67 resulted in an increase of about three orders of magnitude in the specificity towards α-D-galactopyranoside, with an only 2.7-fold decrease in the specificity towards β-L-arabinopyranoside. These results confirm the general structure-specificity arguments presented above, and explain the preference of Abp for arabino substrates compared with galacto substrates. Apparently, despite the significant increase in the specific activity towards galactose substrates, a single mutation was not sufficient to introduce a new specificity to Abp comparable to the original specificity of the WT enzyme towards arabinose substrates. Additional mutations around the active site would probably be required for such a specificity change, mutations that, in principle, can be deduced from homologous bifunctional enzymes. Candidates for such deduction could be, for example, the two GH27 enzymes (Fo/AP1 and Fo/AP2) recently isolated from *Fusarium*

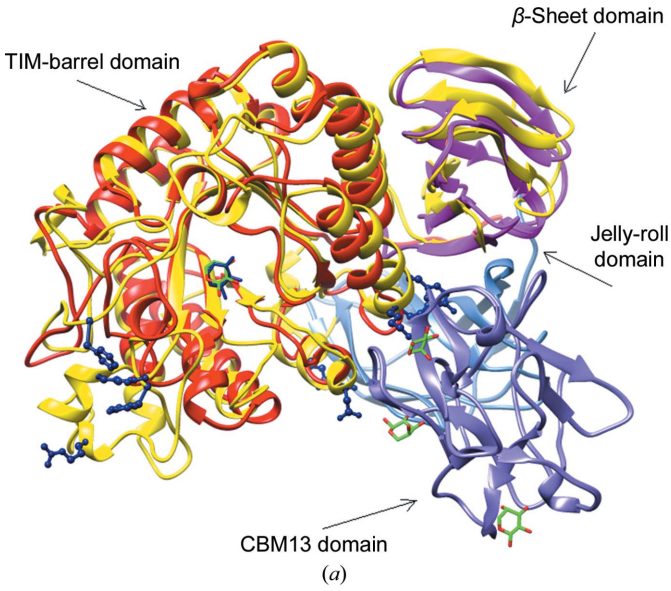
Figure 9
Structural comparison of Abp with 3cc1, a highly homologous protein (PDB entry 3cc1). (*a*) A superposition of the Abp basic dimer (chains *A* and *C*) with the 3cc1 dimer present in the asymmetric unit. The Abp dimer is shown in red and orange, while the 3cc1 dimer is shown in blue and green. A very good fit is demonstrated for a monomer overlap of the two structures (the red and blue monomers); however, the dimer arrangements differ significantly. (*b*) A comparison of the Abp tetramer (right) with the 3cc1 tetramer (left) as generated with the *PISA* server. It is shown that the Abp tetramer adopts a significantly more 'open' conformation relative to 3cc1. (*c*) A superposition of the active site of Abp-WT (green; labelled in black) on the active site of 3cc1 (purple; labelled in red). The residues involved in substrate binding are shown in bright green for Abp and in magenta for 3cc1, demonstrating both similarities and differences between the corresponding residues. The catalytic residues and the specificity-governing Ile residues are shown in yellow.

oxysporum 12S, which show comparable β-L-arabinopyranosidase and α-D-galactopyranosidase activities (Sakamoto *et al.*, 2010).

3.5. Comparison with related protein structures

3.5.1. Comparison with 3cc1, a highly homologous protein. Among the enzymes of the GH27 family for which three-dimensional structures are available, the protein that shows the highest homology to Abp is a putative α-*N*-acetylgalactosaminidase (BH1870) from *B. halodurans* C-125 (PDB entry 3cc1; Joint Center for Structural Genomics, unpublished work). The structure of this protein (referred to here as 3cc1) was also used as the reference model for the molecular-replacement calculations that led to the structure determination of Abp (see above). The 3cc1 protein contains 433 residues and shares 65% sequence identity with Abp. Superposition of the monomers of Abp-WT and 3cc1 demonstrates a very good fit (Fig. 9*a*), as reflected in an r.m.s.d. value of 0.96 Å calculated over 387 aligned residues.

However, although the tertiary structures of Abp and 3cc1 match very well, their quaternary structures do not. The crystallographic asymmetric unit of 3cc1 contains a dimer, but according to the information reported in the corresponding PDB entry 3cc1 is a tetramer in solution. Superposing the 3cc1 asymmetric dimer upon the Abp basic dimer (chains *A* and *C*), it is apparent that although both dimers resemble 'pincers', the Abp pincers are significantly more open than the 3cc1 pincers (Fig. 9*a*). Using the *PISA* server (Krissinel & Henrick, 2007) it was possible to generate a 3cc1 'tetramer'. A comparison between the Abp and 3cc1 tetramers (Fig. 9*b*) demonstrates that although the Abp and 3cc1 tetramers are both arranged with a similar oligomeric architecture, the overall quaternary structure is significantly different. Both proteins are composed



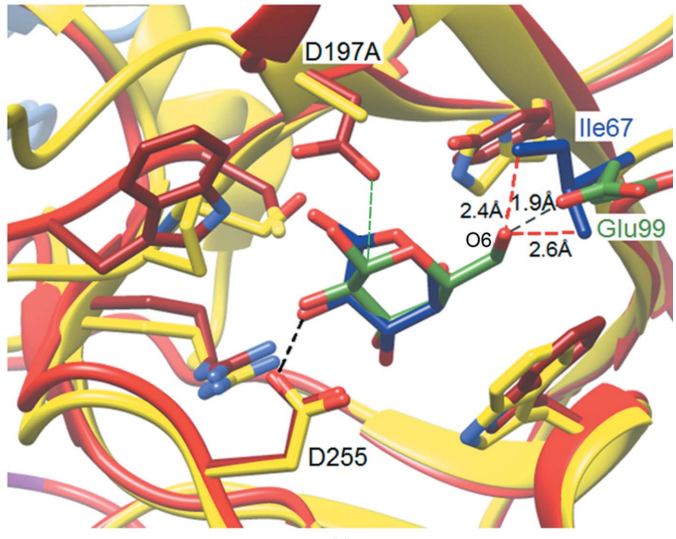


Figure 10
A comparison of the current structure of Abp with the structure of SaArap27A (PDB entries 3a21, 3a22 and 3a23), a related arabinopyranosidase from the GH27 family. (*a*) A superposition of the monomers of Abp (chain *A*, yellow) and SaArap27A (chain *B*, other colours). The four domains of SaArap27A, the TIM-barrel domain, the β-sheet domain, the CBM13 domain and the jelly-roll domain, are shown in red, pink, purple and blue, respectively. The comparison demonstrates the relatively high structural similarities between the two Abp domains and the corresponding TIM-barrel and β-sheet domains of SaArap27A. Small differences are observed in the Abp monomer parts that interact with neighbouring monomers forming the Abp tetramer. The residues participating in these contacts are shown in dark blue (ball-and-stick representation). The captured arabinose molecules are also shown (blue for Abp; green for SaArap27A). (*b*) Electrostatic surface potential of the SaArap27A monomer (calculated with the *APBS* plug-in as implemented in *PyMOL*). The potential is calculated according to the linearized Poisson–Boltzmann equation using the PARSE force field at pH 7. The potential gradient is in the range $3kT$/e to $-3kT$/e, where dark blue represents the most positive potential and dark red represents the most negative potential. Lower potential levels are shown in lighter shades of the corresponding colour. A local concentration of negative potential is noticed between the catalytic TIM-barrel domain and the CBM13 domain. Bound arabinose molecules are shown in green. (*c*) A superposition of the active sites of Abp-D197A–ARB (in yellow) and SaArap27A (in red). Most of the residues involved in binding the sugar product molecules are shown, demonstrating relatively high similarities in the product-binding modes of the two enzymes. A pronounced difference, however, is caused by the replacement of Ile67 in Abp (blue) by the corresponding Glu99 in SaArap27A (green). Bound sugar products are shown in blue (Abp) and green (SaArap27A).

of a dimers of dimers, and in both proteins the two dimers are clamped around each other to form 'staggered' tetrameric structures. However, in accordance with the differences observed for the dimers, the 3cc1 tetramer is significantly more 'closed' than the Abp tetramer, making it more compact and making its central cavity much narrower. These differences in quaternary conformation may suggest that the two proteins can adopt two different conformational states, an 'open' state and a 'closed' state, and that these two states may interconvert into one another for various stability and/or functional purposes (for example to allow substrate entrance into the central cavity). These differences may also reflect the differences in the size and nature of the specific substrates of the two enzymes, either directly or indirectly.

The active sites of Abp and 3cc1 are similar, in accordance with the high sequence homology between the proteins (Fig. 9*c*). Moreover, the 3cc1 active site contains the residue Ile53, which is analogous to Ile67 in Abp. As stated above, this residue is thought to govern specificity towards arabinose in the specific case of Abp, taking into account that in the structures of most α-*N*-acetylgalactosaminidases and galactosidases the analogous residue in this position is Asp. Since Asp and Ile are quite different in terms of both size, shape and chemical properties, they lead to different binding interactions with the substrate (see above). In light of this similarity of 3cc1 to Abp and the difference in this position in other α-*N*-acetylgalactosaminidases, it seems as if the biological function of 3cc1 may not be that of an α-*N*-acetylgalactosaminidase, as stated in the PDB information file, but rather that of an arabinopyranosidase, similar to Abp. This is also in a better accord with the very high homology (in terms of both structure and sequence) between Abp and 3cc1.

Nevertheless, a few differences exist as well. For one thing, the active-site residues Ser231, Pro232 and Arg251 of Abp are positioned quite differently compared with the analogous residues in 3cc1, which are Ser217, Pro218 and Arg237 (Fig. 9*c*). Ser217 and Pro218 in 3cc1 are situated on a loop further away from the analogous loop in Abp, apparently making the binding site of 3cc1 slightly larger than that of Abp, and the loop on which Arg237 in 3cc1 is situated is in a completely different position compared with the loop on which Arg251 in Abp is situated (Fig. 9*c*). It seems as if these differences could be related to the differences in the true natural substrates of the two enzymes, as these residues are involved in determining the exact space and shape of the substrate-binding region, specifically making the binding site of Abp smaller than the corresponding binding site of 3cc1. We thus see that despite the similarity in the key Ile residue in both enzymes, the true substrates of the two proteins may be slightly different. This should be especially relevant for the parts that constitute the reducing part of the sugar substrate, which bind to the +*n* subsites. Obviously, however, answers to this question await further experimental data.

3.5.2. Comparison with a related GH27 arabinopyranosidase. To the best of our knowledge, only one other GH27 enzyme with confirmed β-L-arabinopyranosidase activity has been structurally characterized to date. This is the SaArap27A protein isolated from *Streptomyces avermitilis* (PDB entries 3a21, 3a22 and 3a23; Ichinose *et al.*, 2009), which shares 23% sequence identity with Abp. SaArap27A is significantly larger than Abp, contains 614 amino acids compared with the 448 in Abp and possesses four different domains compared with the two domains of Abp (Fig. 10*a*). The four domains of SaArap27A include a catalytic TIM-barrel domain, a β-sheet domain, a jelly-roll structure domain and a carbohydrate-binding module family 13 (CBM13) domain. Of these domains, the first two are also present in Abp and are characteristic of the entire GH27 family, as discussed above.

A superposition of the monomers of Abp and SaArap27A gives a reasonable fit over the two shared domains (Fig. 10*a*), with an r.m.s.d. value of 1.86 Å calculated over 330 common residues. However, within the shared domains (the TIM-barrel and the all-β domains) there are still regions that differ significantly between the two proteins. One such region is the area in Abp that contains the residues forming the interface within the Abp basic dimers (interface type I between chains *A* and *C* and chains *B* and *D*; see §3.1.2 above). As seen in Fig. 10(*a*), this region is completely missing in SaArap27A. In addition, there are no residues in SaArap27A that are analogous to the residues forming the interface between the Abp dimers (interface type II between chains *A* and *C* and chains *B* and *D*). Considering the absence of these contact-forming residues in SaArap27A, it is not surprising that SaArap27A functions as a monomer (Ichinose *et al.*, 2009) and not as a tetramer, as is the case for Abp. The reason for these oligomerization differences is as yet unclear, but could be related, directly or indirectly, to the presence of the two extra domains in SaArap27A. In this respect, it is interesting to note that in SaArap27A the extra CBM13 domain is situated in the general location of interface II in Abp (Fig. 10*a*), indicating that this extra domain in SaArap27A prevents the intermolecular interactions observed in this region in Abp.

Moreover, looking at the electrostatic surface potential of SaArap27A, it is apparent that a significant negative potential is concentrated between the catalytic TIM-barrel domain and the CBM13 extra domain (Fig. 10*b*). In light of the enhanced negative electrostatic potential caused by the oligomerization of Abp (see above), it seems as if a similar electrostatic enhancement is enabled by the extra domains in SaArap27A, possibly to help attract substrate to the protein by a synergistic accumulation of negative charge around the general area of the active site. Interestingly, and probably regardless of the possible electrostatic effect of the CBM13 domain in SaArap27A (Fig. 10*b*), this domain was found to bind a few additional L-arabinose molecules (Ichinose *et al.*, 2009). These extra arabinose sites (Figs. 10*a* and 10*b*) indicate alternative substrate binding and probably reflect the original functional role of this domain in SaArap27A and related proteins.

As expected, the active sites of Abp and SaArap27A are quite similar, as demonstrated in Fig. 10(*c*). A noticeable difference, however, concerns residue Glu99 in SaArap27A, which is the analogous residue to Ile67 in Abp. Previous work on SaArap27A has shown that similarly to Abp, SaArap27A is

considerably more active towards PNP-β-L-AraP compared with pND-α-D-GalP, which indicates that the preferred substrates of this enzyme should have an arabinose moiety rather than a galactose moiety at the nonreducing side of the enzymatic bond cleavage. Interestingly, when Glu99 was changed to Asp in this position, this substrate preference was reversed (galactose substrates were preferred over arabinose substrates), suggesting that Glu99 is a key element in determining the substrate specificity of SaArap27A (Ichinose *et al.*, 2009). The explanation given by the authors was that there is a tight binding of the galactose molecule to the active site of native SaArap27A resulting from the strong hydrogen bond between Glu99 and the O6 atom of the bound galactose. It was claimed that such tight binding reduces the turnover of the catalysis, so that when it is substituted by Asp the binding decreases, the bound galactose can be released more easily and the overall catalytic turnover improved (Ichinose *et al.*, 2009).

Nevertheless, when we closely checked the specific enzyme–galactose interactions in the reported stucture of SaArap27A–galactose complex (PDB entry 3a23), it appeared that the distance between Glu99 and the target O6 atom of the bound galactose is too short, even for a 'tight hydrogen bond' as claimed. These distances are about 1.1 and 1.9 Å for monomers *A* and *B* of SaArap27A, respectively, compared with 'normal' hydrogen bonds that are usually in the range 2.4–3.0 Å, raising questions about the reliability of the crystallographic model building in this specific region. This is especially true for the validity of the structural modelling and the exact positioning of the bound galactose molecule in this structure. Rather, it seems that in a similar manner to Ile67 in Abp (see above), Glu99 would form a clash with O6 of a galactose molecule if bound in this site, and this better explains the low activity of the wild-type SaArap27A towards galactose substrates. We therefore suggest that although the original specificity-governing residue differs significantly in Abp and SaArap27A in terms of its chemical properties (hydrophobic Ile67 in Abp *versus* negatively charged Glu99 in SaArap27A), in both proteins this residue influences the catalytic activity in an analogous manner towards similar substrate specificities. In this respect, the increase in k_{cat}/K_m towards galactose substrates observed for the Abp-I67D mutant (about three orders of magnitude) is significantly larger than the relatively moderate effect of the analogous mutation in the SaArap27A-E99D mutant (only about one order of magnitude; Ichinose *et al.*, 2009). Such a difference could probably be accounted for by the more radical change in the side chain in the case of Abp (Ile to Asp) compared with the less radical change in the case of SaArap27A (Glu to Asp). Obviously, it will be possible to confirm the generality of these observations and their suggested structure–function interpretations as more structures of GH27 β-L-arabinopyranosidases with different specificities are determined and reported.

The research leading to the results presented here received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under BioStruct-X (grant agreement No. 283570). This work was also supported by the Israel Science Foundation Grants 500/10 and 152/11, the I-CORE Program of the Planning and Budgeting Committee, the Ministry of Environmental Protection and the Grand Technion Energy Program (GTEP), and comprises part of The Leona M. and Harry B. Helmsley Charitable Trust Reports on Alternative Energy series of the Technion, Israel Institute of Technology and the Weizmann Institute of Science. YS acknowledges partial support by the Russell Berrie Nanotechnology Institute and The Lorry I. Lokey Interdisciplinary Center for Life Science and Engineering, Technion. We thank the staff at the European Synchrotron Research Facility (ESRF, BM14 beamline) and EMBL for their helpful support in the X-ray synchrotron data measurement and analysis. The synchrotron experiments at ESRF were also supported by the ESRF internal funding program. YS holds the Erwin and Rosl Pollak Chair in Biotechnology at the Technion.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Alalouf, O., Balazs, Y., Volkinshtein, M., Grimpel, Y., Shoham, G. & Shoham, Y. (2011). *J. Biol. Chem.* **286**, 41993–42001.
Alhassid, A., Ben-David, A., Tabachnikov, O., Libster, D., Naveh, E., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2009). *Biochem. J.* **422**, 73–82.
Allaire, M. & Yang, L. (2011). *J. Synchrotron Rad.* **18**, 41–44.
Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. (2010). *Nucleic Acids Res.* **38**, W529–W533.
Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
Bar, M., Golan, G., Nechama, M., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2004). *Acta Cryst.* D**60**, 545–549.
Ben-David, A., Bravman, T., Balazs, Y. S., Czjzek, M., Schomburg, D., Shoham, G. & Shoham, Y. (2007). *Chembiochem*, **8**, 2145–2151.
Ben-David, A., Shoham, G. & Shoham, Y. (2008). *Chem. Biol.* **15**, 546–551.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Böhm, A., Diez, J., Diederichs, K., Welte, W. & Boos, W. (2002). *J. Biol. Chem.* **277**, 3708–3717.
Bravman, T., Belakhov, V., Solomon, D., Shoham, G., Henrissat, B., Baasov, T. & Shoham, Y. (2003). *J. Biol. Chem.* **278**, 26742–26749.
Bravman, T., Mechaly, A., Shulami, S., Belakhov, V., Baasov, T., Shoham, G. & Shoham, Y. (2001). *FEBS Lett.* **495**, 115–119.
Bravman, T., Zolotnitsky, G., Belakhov, V., Shoham, G., Henrissat, B., Baasov, T. & Shoham, Y. (2003). *Biochemistry*, **42**, 10528–10536.
Bravman, T., Zolotnitsky, G., Shulami, S., Belakhov, V., Solomon, D., Baasov, T., Shoham, G. & Shoham, Y. (2001). *FEBS Lett.* **495**, 39–43.
Brüx, C., Ben-David, A., Shallom-Shezifi, D., Leon, M., Niefind, K., Shoham, G., Shoham, Y. & Schomburg, D. (2006). *J. Mol. Biol.* **359**, 97–109.
Brüx, C., Niefind, K., Ben-David, A., Leon, M., Shoham, G., Shoham, Y. & Schomburg, D. (2005). *Acta Cryst.* F**61**, 1054–1057.
Cardoso, S. M., Silva, A. M. & Coimbra, M. A. (2002). *Carbohydr. Res.* **337**, 917–924.
Clark, N. E. & Garman, S. C. (2009). *J. Mol. Biol.* **393**, 435–447.
Cremer, D. & Pople, J. A. (1975). *J. Am. Chem. Soc.* **97**, 1354–1358.
Dann, R., Lansky, S., Lavid, N., Zehavi, A., Belakhov, V., Baasov, T., Duir, H., Manjasetty, B. A., Belrhali, H., Shoham, Y. & Shoham, G. (2014). In the press.

Davies, G. J. & Henrissat, B. (1995). *Structure*, **3**, 853–859.
Davies, G. J., Planas, A. & Rovira, C. (2012). *Acc. Chem. Res.* **45**, 308–316.
Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem. J.* **321**, 557–559.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Ermel, F. F., Follet-Gueye, M.-L., Cibert, C., Vian, B., Morvan, C., Catesson, A.-M. & Goldberg, R. (2000). *Planta*, **210**, 732–740.
Fernández-Leiro, R., Pereira-Rodríguez, A., Cerdán, M. E., Becerra, M. & Sanz-Aparicio, J. (2010). *J. Biol. Chem.* **285**, 28020–28033.
Fujimoto, Z., Kaneko, S., Kim, W.-D., Park, G.-G., Momma, M. & Kobayashi, H. (2009). *Biosci. Biotechnol. Biochem.* **73**, 2360–2364.
Fujimoto, Z., Kaneko, S., Momma, M., Kobayashi, H. & Mizuno, H. (2003). *J. Biol. Chem.* **278**, 20313–20318.
Garman, S. C. & Garboczi, D. N. (2004). *J. Mol. Biol.* **337**, 319–335.
Garman, S. C., Hannick, L., Zhu, A. & Garboczi, D. N. (2002). *Structure*, **10**, 425–434.
Gerber, S., Comellas-Bigler, M., Goetz, B. A. & Locher, K. P. (2008). *Science*, **321**, 246–250.
Gilead, S. & Shoham, Y. (1995). *Appl. Environ. Microbiol.* **61**, 170–174.
Golan, G., Shallom, D., Teplitsky, A., Zaide, G., Shulami, S., Baasov, T., Stojanoff, V., Thompson, A., Shoham, Y. & Shoham, G. (2004). *J. Biol. Chem.* **279**, 3014–3024.
Golubev, A. M., Nagem, R. A. P., Brandão Neto, J. R., Neustroev, K. N., Eneyskaya, E. V., Kulminskaya, A. A., Shabalin, K. A., Savel'ev, A. N. & Polikarpov, I. (2004). *J. Mol. Biol.* **339**, 413–422.
Guce, A. I., Clark, N. E., Salgado, E. N., Ivanen, D. R., Kulminskaya, A. A., Brumer, H. & Garman, S. C. (2010). *J. Biol. Chem.* **285**, 3625–3632.
Guinier, A. (1939). *Ann. Phys. (Paris)*, **12**, 161–237.
Hollenstein, K., Frei, D. C. & Locher, K. P. (2007). *Nature (London)*, **446**, 213–216.
Hövel, K., Shallom, D., Niefind, K., Baasov, T., Shoham, G., Shoham, Y. & Schomburg, D. (2003). *Acta Cryst.* D**59**, 913–915.
Hövel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. & Schomburg, D. (2003). *EMBO J.* **22**, 4922–4932.
Huisman, M. M., Brüll, L. P., Thomas-Oates, J. E., Haverkamp, J., Schols, H. A. & Voragen, A. G. (2001). *Carbohydr. Res.* **330**, 103–114.
Ichinose, H., Fujimoto, Z., Honda, M., Harazono, K., Nishimoto, Y., Uzura, A. & Kaneko, S. (2009). *J. Biol. Chem.* **284**, 25097–25106.
IUPAC–IUB Joint Commission on Biochemical Nomenclature (1980). *Eur. J. Biochem.* **111**, 295–298.
Kadaba, N. S., Kaiser, J. T., Johnson, E., Lee, A. & Rees, D. C. (2008). *Science*, **321**, 250–253.
Knott, B. C., Haddad Momeni, M., Crowley, M. F., Mackenzie, L. F., Götz, A. W., Sandgren, M., Withers, S. G., Ståhlberg, J. & Beckham, G. T. (2014). *J. Am. Chem. Soc.* **136**, 321–329.
Konarev, P. V., Petoukhov, M. V., Volkov, V. V. & Svergun, D. I. (2006). *J. Appl. Cryst.* **39**, 277–286.
Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.
Konishi, T., Takeda, T., Miyazaki, Y., Ohnishi-Kameyama, M., Hayashi, T., O'Neill, M. A. & Ishii, T. (2007). *Glycobiology*, **17**, 345–354.
Kozin, M. B. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 33–41.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Lansky, S., Alalouf, O., Salama, R., Dvir, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* F**70**, 476–481.
Lansky, S., Alalouf, O., Solomon, H. V., Alhassid, A., Govada, L., Chayen, N. E., Belrhali, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* D**70**, 261–278.
Lansky, S., Alalouf, O., Solomon, V., Alhassid, A., Govada, L., Chayen, N. E., Belrhali, H., Shoham, Y. & Shoham, G. (2013). *Acta Cryst.* F**69**, 430–434.
Lansky, S., Salama, R., Dann, R., Shner, I., Manjasetty, B. A., Belrhali, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* F**70**, 1038–1045.
Lansky, S., Salama, R., Solomon, V. H., Belrhali, H., Shoham, Y. & Shoham, G. (2013). *Acta Cryst.* F**69**, 695–699.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
McCarter, J. D. & Withers, S. G. (1994). *Curr. Opin. Struct. Biol.* **4**, 885–892.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Mohnen, D. (2008). *Curr. Opin. Plant Biol.* **11**, 266–277.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). *J. Appl. Cryst.* **45**, 342–350.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Pinkett, H. W., Lee, A. T., Lum, P., Locher, K. P. & Rees, D. C. (2007). *Science*, **315**, 373–377.
Ponder, G. R. & Richards, G. N. (1997). *Carbohydr. Polym.* **34**, 251–261.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Rees, D. C., Johnson, E. & Lewinson, O. (2009). *Nature Rev. Mol. Cell Biol.* **10**, 218–227.
Sakamoto, T., Tsujitani, Y., Fukamachi, K., Taniguchi, Y. & Ihara, H. (2010). *Appl. Microbiol. Biotechnol.* **86**, 1115–1124.
Salama, R., Alalouf, O., Tabachnikov, O., Zolotnitsky, G., Shoham, G. & Shoham, Y. (2012). *FEBS Lett.* **586**, 2436–2442.
Shallom, D., Belakhov, V., Solomon, D., Gilead-Gropper, S., Baasov, T., Shoham, G. & Shoham, Y. (2002). *FEBS Lett.* **514**, 163–167.
Shallom, D., Belakhov, V., Solomon, D., Shoham, G., Baasov, T. & Shoham, Y. (2002). *J. Biol. Chem.* **277**, 43667–43673.
Shallom, D., Golan, G., Shoham, G. & Shoham, Y. (2004). *J. Bacteriol.* **186**, 6928–6937.
Shallom, D., Leon, M., Bravman, T., Ben-David, A., Zaide, G., Belakhov, V., Shoham, G., Schomburg, D., Baasov, T. & Shoham, Y. (2005). *Biochemistry*, **44**, 387–397.
Shofiqur Rahman, A. K. M., Kato, K., Kawai, S. & Takamizawa, K. (2003). *Carbohydr. Res.* **338**, 1469–1476.
Shulami, S., Gat, O., Sonenshein, A. L. & Shoham, Y. (1999). *J. Bacteriol.* **181**, 3695–3704.
Shulami, S., Raz-Pasteur, A., Tabachnikov, O., Gilead-Gropper, S., Shner, I. & Shoham, Y. (2011). *J. Bacteriol.* **193**, 2838–2850.
Shulami, S., Zaide, G., Zolotnitsky, G., Langut, Y., Feld, G., Sonenshein, A. L. & Shoham, Y. (2007). *Appl. Environ. Microbiol.* **73**, 874–884.
Solomon, V., Teplitsky, A., Shulami, S., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2007). *Acta Cryst.* D**63**, 845–859.
Stawiski, E. W., Mandel-Gutfreund, Y., Lowenthal, A. C. & Gregoret, L. M. (2002). *Pac. Symp. Biocomput.*, pp. 637–648.
Svergun, D. (1999). *Biophys. J.* **76**, 2879–2886.
Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Tabachnikov, O. & Shoham, Y. (2013). *FEBS J.* **280**, 950–964.
Teplitsky, A., Feinberg, H., Gilboa, R., Lapidot, A., Mechaly, A., Stojanoff, V., Capel, M., Shoham, Y. & Shoham, G. (1997). *Acta Cryst.* D**53**, 608–611.
Teplitsky, A., Mechaly, A., Stojanoff, V., Sainz, G., Golan, G., Feinberg, H., Gilboa, R., Reiland, V., Zolotnitsky, G., Shallom, D.,

Thompson, A., Shoham, Y. & Shoham, G. (2004). *Acta Cryst.* D**60**, 836–848.
Teplitsky, A., Shulami, S., Moryles, S., Shoham, Y. & Shoham, G. (2000). *Acta Cryst.* D**56**, 181–184.
Teplitsky, A., Shulami, S., Moryles, S., Zaide, G., Shoham, Y. & Shoham, G. (1999). *Acta Cryst.* D**55**, 869–872.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.
Ward, A., Reyes, C. L., Yu, J., Roth, C. B. & Chang, G. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 19005–19010.
Willför, S., Sjöholm, R., Laine, C. & Holmbom, B. (2002). *Wood Sci. Technol.* **36**, 101–110.
Yang, L. (2013). *J. Synchrotron Rad.* **20**, 211–218.
Zaide, G., Shallom, D., Shulami, S., Zolotnitsky, G., Golan, G., Baasov, T., Shoham, G. & Shoham, Y. (2001). *Eur. J. Biochem.* **268**, 3006–3016.