Preliminary crystallographic analysis of Xyn52B2, a GH52 β -D-xylosidase from *Geobacillus stearothermophilus* T6

Roie Dann,^a Shifra Lansky,^a
Noa Lavid,^b Arie Zehavi,^b Valery
Belakhov,^c Timor Baasov,^c
Hay Dvir,^d Babu Manjasetty,^{e,f}
Hassan Belrhali,^{e,f} Yuval
Shoham^{b*} and Gil Shoham^{a*}

^aInstitute of Chemistry and the Laboratory for Structural Chemistry and Biology, The Hebrew University of Jerusalem, Jerusalem 91904, Israel,

^bDepartment of Biotechnology and Food Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel, ^cSchulich Faculty of Chemistry, Technion – Israel Institute of Technology, Haifa 32000, Israel, ^dTechnion Center for Structural Biology, Lorry I. Lokey Center for Life Sciences and Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel, ^eEuropean Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, 38000 Grenoble, France, and ^fUnit 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 29 August 2014
Accepted 29 October 2014

© 2014 International Union of Crystallography
All rights reserved

Geobacillus stearothermophilus T6 is a thermophilic bacterium that possesses an extensive hemicellulolytic system, including over 40 specific genes that are dedicated to this purpose. For the utilization of xylan, the bacterium uses an extracellular xylanase which degrades xylan to decorated xylo-oligomers that are imported into the cell. These oligomers are hydrolyzed by side-chain-cleaving enzymes such as arabinofuranosidases, acetylsterases and a glucuronidase, and finally by an intracellular xylanase and a number of β -xylosidases. One of these β -xylosidases is Xyn52B2, a GH52 enzyme that has already proved to be useful for various glycosynthesis applications. In addition to its demonstrated glycosynthase properties, interest in the structural aspects of Xyn52B2 stems from its special glycoside hydrolase family, GH52, the structures and mechanisms of which are only starting to be resolved. Here, the cloning, overexpression, purification and crystallization of Xyn52B2 are reported. The most suitable crystal form that has been obtained belonged to the orthorhombic $P2_12_12_1$ space group, with average unit-cell parameters $a = 97.7$, $b = 119.1$, $c = 242.3$ Å. Several X-ray diffraction data sets have been collected from flash-cooled crystals of this form, including the wild-type enzyme (3.70 Å resolution), the E335G catalytic mutant (2.95 Å resolution), a potential mercury derivative (2.15 Å resolution) and a selenomethionine derivative (3.90 Å resolution). These data are currently being used for detailed three-dimensional structure determination of the Xyn52B2 protein.

1. Introduction

Intensive research efforts have been devoted in recent years to biotechnologically relevant thermostable enzymes, especially those originating from thermophilic bacteria. One of these target bacteria is *Geobacillus stearothermophilus* T6, a Gram-positive bacterium that is widely distributed in soil and thermophilic habitats such as thermal vents and ocean sediments. The special thermophilic features of these bacteria allow them to withstand relatively high temperatures, sometimes over 75°C (Donk, 1920). This bacterium thrives on plant cell-wall oligosaccharides, including xylan, arabinan and galactan (Shulami *et al.*, 1999, 2007, 2011, 2014; Tabachnikov & Shoham, 2013), for which it possesses comprehensive enzymatic utilization systems. Each of these integrated systems includes a battery of specialized enzymes, most of which are glycoside hydrolases (GHs) that work synergistically to degrade the environmental oligosaccharides into smaller subunits. These subunits are then used by the bacterium for a wide range of cellular processes. The best characterized of these complex utilization systems is that of xylan, where the bacterium secretes a specific extracellular xylanase (XT6; Gat *et al.*, 1994; Teplitsky *et al.*, 1997, 2004; Bar *et al.*, 2004) that breaks down the polymeric xylan into shorter xylo-oligosaccharides. These short, usually decorated, sugar oligomers are then transported into the bacterial cell *via* specific ABC sugar transporters (Shulami *et al.*, 2007; Rees *et al.*, 2009). Inside the cell, the decorated xylo-oligosaccharides are hydrolyzed by several 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, Shallom *et al.*, 2004; Shallom *et al.*,

2004), acetylsterases (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 (IXT6; Teplitzky *et al.*, 2000; Solomon *et al.*, 2007) and several xylosidases (Bravman *et al.*, 2001, 2003; Shallom *et al.*, 2005; Brx *et al.*, 2006; Ben-David *et al.*, 2007). The xylosidases perform the last step, hydrolyzing the short linear xylo-oligomers [usually xylotriose (X3) and xylobiose (X2)] into xylose monomers. One of these key xylosidases of *G. stearothermophilus* T6 is Xyn52B2, the subject of the current study.

Xylanases and xylosidases from thermophilic bacteria have proven to have a wide range of potential industrial and biotechnological applications, for example their use as an environmentally friendly alternative to chlorine dioxide in large-scale pulp-bleaching processes (Shoham *et al.*, 1992; Lundgren *et al.*, 1994; Suurnkki *et al.*, 1997). Such applications also include the improvement of the digestibility of animal feed, the production of D-xylose for xylitol manufacture, the breakdown of lignocellulosic waste to produce biofuel (Yeoman *et al.*, 2010) and the enzymatic production of oligosaccharides from simpler sugar units as enabled by their conversion from glycosidases to glycosynthases (Mackenzie *et al.*, 1998; Hancock *et al.*, 2006). Glycosynthases are mutants of glycoside hydrolases in which the acidic catalytic nucleophile is replaced by a small non-nucleophilic residue, resulting in a practically inactive hydrolase. However, in the presence of glycosyl fluorides of the opposite anomeric configuration (to their substrates), these enzymes completely change their nature, and transfer the activated sugars to suitable acceptors (Ben-David *et al.*, 2007, 2008). The use of robust enzymes that can accommodate non-natural substrates, and the use of inexpensive glycosyl-fluoride donors, make such a glycosynthase approach a very attractive option for efficient oligosaccharide synthesis.

The GH52 β -xylosidase from *G. stearothermophilus* T6 (EC 3.2.1.37), Xyn52B2, consists of 705 amino acids with a calculated molecular weight of 79 894. The enzyme has been characterized biochemically (Bravman *et al.*, 2001, 2003; Jordan *et al.*, 2013) and has been shown to hydrolyze its substrate with a net retention of the anomeric configuration (Bravman *et al.*, 2001). The substrate specificity for the glycon moiety of the enzyme is considerably restricted to xylose, and binding of xylobiose and xylotriose was shown to be enthalpy-driven. Catalysis appeared to depend on two ionizable amino-acid residues: the acid-base (Asp495) and the nucleophile (Glu335). The Brnsted plot relationship was consistent with a double-displacement mechanism, as expected for retaining enzymes, involving the formation and breakdown of an enzyme–substrate intermediate (Bravman *et al.*, 2003). Xyn52B2 was shown to be converted into a significant glycosynthase by site-directed mutagenesis of the catalytic nucleophile (*e.g.* E335G) and its glycosynthase activity was further improved by a few rounds of directed evolution (Ben-David *et al.*, 2007, 2008). The enzyme could also be modified by a single amino-acid replacement (Y509E) in order to obtain xylanase activity in addition to the original xylosidase activity (Huang *et al.*, 2014).

Previous experiments to crystallize Xyn52B2 for structural analysis resulted in a triclinic crystal form (space group *P*1), which gave reasonable X-ray diffraction to 2.0 resolution (Czjzek *et al.*, 2004). Unfortunately, efforts to reproduce these or similar crystals for further crystallographic analysis proved unsuccessful for a long time, the reasons for which remain unclear. Recently, it was decided to initiate a new series of crystallization experiments in order to obtain suitable crystals for full structure determination. In the present report we describe these recent crystallization experiments, which resulted in the preliminary crystallographic characterization of wild-type Xyn52B2 (Xyn52B2-WT) as well as its nucleophile catalytic mutant

Xyn52B2-E335G. We also describe the production of and the single-wavelength anomalous diffraction (SAD) data collection from the fully substituted selenomethionine derivative of the enzyme (Xyn52B2-E335G-Se), which is expected to lead to the detailed structure determination of this protein. As only one other structure of a GH52 protein has currently been reported (Espina *et al.*, 2014), the structure of Xyn52B2 should be highly important not only for understanding the structure and mechanism of this particular enzyme, but also for a more general understanding of the structure–activity relationships within the GH52 family.

2. Experimental

2.1. Overexpression and purification of Xyn52B2-WT and Xyn52B2-E335G

Overexpression and purification of Xyn52B2-WT and the nucleophile catalytic mutant Xyn52B2-E335G were essentially as described previously (Bravman *et al.*, 2001, 2003; Ben-David *et al.*, 2007, 2008). Briefly, the *xynB2* gene (GenBank accession No. AJ305327) was cloned in the pET9d vector and overexpressed in *Escherichia coli* BL21(DE3) cells. The purification procedure included two steps, a heat-treatment step at 333 K followed by a gel-filtration step, resulting in gram quantities of the protein with over 99% purity (Bravman *et al.*, 2001).

2.2. Production and purification of Xyn52B2-Se

Fully substituted selenomethionine-derivatized protein (Xyn52B2-E335G-Se) was prepared on an *E. coli* B834(DE3) (Met[−]) background, following the general procedure described previously (Mechaly *et al.*, 2000). Briefly, the culture was grown on defined medium consisting of 2 g l^{−1} NH₄Cl, 6 g l^{−1} KH₂PO₄, 12 g l^{−1} Na₂HPO₄, 0.5 g l^{−1} MgSO₄·7H₂O and 0.025 g l^{−1} FeSO₄ supplemented with vitamins (riboflavin, pyridoxine and thiamine; 1 g ml^{−1} each), 5 g l^{−1} glucose, amino acids (Asn, Asp, Cys, Glu, Ala, Arg, Val, Gln, Gly, His, Ile, Leu, Phe, Trp, Tyr, Lys, Pro, Ser and Thr; 50 g ml^{−1} each), 50 g ml^{−1} seleno-L-methionine and 25 g ml^{−1} kanamycin. The starter culture was first grown in 50 ml medium to an OD_{600 nm} of ~4.5 and then transferred into 450 ml fresh medium containing 0.4 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) and grown to a final OD_{600 nm} of ~5. The purification process was the same as for the wild-type protein.

2.3. Preparation of the D-xylose-Hg derivative

A series of thiomeric derivatives of monosaccharides and disaccharides, in which alkylmercury is covalently attached to anomeric thioglycosides, have been specially synthesized to be used as inhibitors and heavy-atom markers for the corresponding xylanases and xylosidases. The synthetic and characterization details of these compounds have been described elsewhere (Belakhov *et al.*, 2001). One of these derivatives, 1-(thiomethylmercuric)- β -D-xyloside (referred to here as X-S-Hg-CH₃), has been used to derivatize Xyn52B2 with mercury for structure-determination purposes. This compound was added as a solid dry powder to the cryoprotection solution used for the flash-cooling procedure of the protein crystals (see below), in a final X-S-Hg-CH₃ concentration of 10 mM. The selected crystals were soaked in this solution for about 30–120 s to allow binding of the compound within the crystal. A fluorescence scan validated the presence of mercury atoms in the soaked crystals, however such a scan is not sufficient to confirm that this is a crystallographically useful Hg-derivative, as the Hg-signal can also

originate from residual mercury atoms in the trapped mother liquor surrounding the crystal.

2.4. Crystallization experiments

Crystallization experiments were set up immediately after the last purification step of the Xyn52B2-WT or Xyn52B2-E335G proteins. In each case, the purified protein was concentrated using Centricon centrifugal concentrators (Millipore, Massachusetts, USA) so that the resulting protein solution consisted of approximately $2\text{--}5\text{ mg ml}^{-1}$ protein, 50 mM Tris-HCl pH 7.0, 100 mM NaCl, 0.02% NaN_3 . This protein solution was used directly for the initial crystallization experiments, which were performed by the sitting-drop and hanging-drop vapour-diffusion methods, using an extensive series of different factorial screening solutions (screens) at a constant temperature of 293 K . The first round of crystallization experiments were set up manually in 4×6 Cryschem hanging-drop crystallization plates (Hampton Research, California, USA), using a number of commercially available sets of screens. These experiments were generally carried out with $5\text{ }\mu\text{l}$ crystallization drops consisting of $2.5\text{ }\mu\text{l}$ protein solution and $2.5\text{ }\mu\text{l}$ reservoir solution, suspended over $600\text{--}1000\text{ }\mu\text{l}$ reservoir solution. Additional rounds of screens were set up in 96-3 iQ TTP Labtech sitting-drop plates using a Mosquito LCP crystal-

lization robot (TTP Labtech) with $0.30\text{ }\mu\text{l}$ crystallization drops consisting of $0.15\text{ }\mu\text{l}$ protein solution and $0.15\text{ }\mu\text{l}$ reservoir solution suspended over $55\text{ }\mu\text{l}$ reservoir solution. The robot-driven crystallization screening experiments were performed at the Technion Center for Structural Biology (TCSB), Technion, Haifa, Israel, using our in-house macromolecular crystallographic setup. These initial screening experiments were used to identify general favourable crystallization conditions for subsequent use as starting points for further refinement (Almog *et al.*, 1993, 1994; Teplitsky *et al.*, 1997, 1999, 2000; Gilboa *et al.*, 1998). When positive crystallization results were observed (*i.e.* crystals or microcrystals), these conditions were repeated with specially constructed home-made solutions, fine-tuning various crystallization parameters such as pH, ionic strength, protein concentration, temperature, precipitating additives, protein drop volume and drop-to-reservoir ratio (Bar *et al.*, 2004; Golan, Zharkov *et al.*, 2004; Reiland *et al.*, 2004; Lansky, Alalouf *et al.*, 2013; Lansky, Salama *et al.*, 2013; Lansky, Alalouf, Salama *et al.*, 2014; Lansky, Salama *et al.*, 2014; Lansky, Zehavi *et al.*, 2014; Solomon *et al.*, 2013). These screening experiments resulted in three different crystal forms (see below), one of which was later found to be suitable for further crystallographic analysis.

Most of the X-ray diffraction data measurements were performed at the European Synchrotron Research Facility (ESRF), Grenoble,

Figure 1
Typical crystals of Xyn52B2-E335G. (a) A complete $5\text{ }\mu\text{l}$ crystallization drop from a typical hanging-drop experiment, showing fully grown crystals of the ESF1 crystal form. The length of the observed crystals is in the range $0.05\text{--}0.3\text{ mm}$. (b) A close-up view of a typical 'elongated snowflake' cluster of the ESF1 crystals, demonstrating the unusual shapes of the crystals and their clusters. (c) The $0.30\text{ }\mu\text{l}$ crystallization micro-drop produced by the Mosquito LCP crystallization robot which resulted in the ERB1 crystal form. (d) The $0.30\text{ }\mu\text{l}$ crystallization micro-drop produced by the Mosquito LCP crystallization robot which resulted in the ERB2 crystal form.

Table 1

Representative parameters for data collection from ESF1-form crystals of Xyn52B2.

Values in parentheses are for the outer shell.

Protein	Xyn52B2-E335G	Xyn52B2-WT	Xyn52B2-E335G-Hg	Xyn52B2-E335G-Se
Diffraction source	TCSB, Technion	BM14, ESRF	BM14, ESRF	BM14, ESRF
Wavelength (Å)	1.541	0.979	0.978	0.978
Temperature (K)	100	100	100	100
Detector	R-AXIS HTC	MAR 225 CCD	MAR 225 CCD	MAR 225 CCD
Crystal-to-detector distance (mm)	230	328	268	328
Rotation range per image (°)	0.5	1.0	0.5	0.5
Total rotation range (°)	220	230	234	288
Exposure time per image (s)	180	13	5	7
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)				
<i>a</i>	97.7	97.5	97.4	97.5
<i>b</i>	119.1	120.0	119.1	119.7
<i>c</i>	242.3	238.3	243.0	237.9
Mosaicity (°)	0.89	0.87	0.44	0.87
Resolution range (Å)	35.00–2.95 (3.00–2.95)	50.00–3.70 (3.90–3.70)	40.00–2.15 (2.19–2.15)	30.00–3.90 (3.97–3.90)
Total No. of reflections	349164 (17042)	256635 (36184)	1240401 (23035)	290688 (14376)
No. of unique reflections	58469 (2839)	30664 (4414)	151718 (6380)	26170 (1284)
Completeness (%)	96.9 (95.5)	100.0 (100.0)	98.5 (85.2)	100.0 (99.9)
Multiplicity	6.0 (6.0)	8.4 (8.2)	8.2 (3.6)	11.1 (11.2)
$\langle I/\sigma(I) \rangle$	6.2 (2.9)	10.7 (3.2)	13.0 (2.9)	5.6 (4.2)
$R_{\text{merge}}^{\dagger}$ (%)	15.4 (61.2)	16.6 (63.0)	10.7 (50.9)	16.3 (63.3)

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

France. Some of the crystals were analyzed at the TCSB, Technion, Haifa, Israel, using our in-house X-ray source (see below). Processing, reduction, indexing, integration and scaling of the ESRF diffraction data were conducted using the *HKL-2000* program suite (Otwinowski & Minor, 1997) and *iMosflm* (Battye *et al.*, 2011). The corresponding processing and scaling of the TCSB diffraction data was performed using the *HKL-3000* program suite (Minor *et al.*, 2006).

3. Results and discussion

3.1. Crystal forms of Xyn52B2-WT and Xyn52B2-E335G

Owing to various technical reasons, most of the initial screening experiments were performed with the nucleophile catalytic mutant of the enzyme, Xyn52B2-E335G. It is therefore these experiments that will be described in detail below. Some of these experiments were also reproduced with Xyn52B2-WT, which generally gave similar results. The crystallization experiments produced three different crystal forms of Xyn52B2-E335G, the first obtained from the initial manually set up screens and the other two obtained at a later stage from the robotically set up screens.

The first crystal form of Xyn52B2-E335G resulted from protein solutions containing 4.7 mg ml^{−1} protein and screening solution No. 18 of the commercial Wizard Classic 2 screen (Emerald Bio, Bainbridge Island, Washington, USA) consisting of 0.2 *M* calcium acetate, 20% PEG 3K, 0.1 *M* Tris buffer pH 7.0. This experimental condition produced quite unusual crystals which looked like ‘elongated snowflakes’ (Figs. 1*a* and 1*b*) and were accordingly termed the ESF1 crystal form. These crystals could be initially observed after 2–5 d of equilibration over 1.0 ml well solutions, and usually appeared as nonsymmetrical clusters built of 3–6 crystals, growing to their final size in an additional 5–10 d. Typically, 2–4 such clusters would appear in the crystallization drop, with the final crystal dimensions of about 0.03 × 0.03 × 0.2 mm (Figs. 1*a* and 1*b*). Despite its very unusual and nonsymmetrical shape, this crystal form was confirmed to represent

single protein crystals and was found suitable for further crystallographic analysis (see below).

The two other crystal forms of Xyn52B2-E335G were obtained from the much smaller scale robotic experiments and resulted from the 4.7 mg ml^{−1} protein solution, with screening solutions Nos. 29 and 39 of the Classics Suite commercial screen (Qiagen, Hilden, Germany). Solution No. 30, consisting of 2.0 *M* ammonium sulfate, 0.1 *M* Tris buffer pH 8.5, produced more usual crystals (compared with the unusual ESF1 crystals), which looked like transparent, well shaped ‘elongated rectangular boxes’ (Fig. 1*c*) and were accordingly termed the ERB1 crystal form. These crystals were initially observed after 14 d of equilibration (about five crystals per drop) and grew to their final size after four additional weeks, with final dimensions of about 0.03 × 0.03 × 0.08 mm (Fig. 1*c*). Solution No. 29, consisting of 2.0 *M* ammonium sulfate, 0.1 *M* sodium acetate buffer pH 4.6, produced similar but significantly smaller crystals termed the ERB2 crystal form (Fig. 1*d*). These crystals were also initially observed after 14 d (about ten crystals per drop) and grew to their final size after six additional weeks, with final dimensions of about 0.02 × 0.02 × 0.06 mm (Fig. 1*d*). Interestingly, the current crystallization conditions for the ERB2 crystal form (of Xyn52B2-E335G) are quite similar to those reported for the original *P1* crystals of Xyn52B2-WT, which were obtained with a well solution consisting of 2.4 *M* ammonium sulfate, 0.04–0.06 *M* sodium citrate buffer pH 5.5 (Czjzek *et al.*, 2004).

After these initial crystals have been obtained, a series of optimization experiments were performed in order to reproduce and improve all three crystal forms. However, of these three crystal forms only the ESF1 crystals could be routinely reproduced, leaving this crystal form the only form that could be used for crystallographic analysis (see below). The two other crystal forms, ERB1 and ERB2, could not be practically reproduced, despite multiple optimization experiments with different purification batches of the protein (wild-type and E335G mutant). The reasons for this behaviour are as yet unclear, but probably reflect the high sensitivity of the Xyn52B2 protein to very small differences in the crystallization solutions, at least with respect to protein conformational stability. Such an abnormal sensitivity could also explain, at least in part, the numerous

unsuccessful attempts to reproduce the original *P1* crystals previously reported for this protein (Czjzek *et al.*, 2004), as mentioned above.

3.2. X-ray diffraction data for Xyn52B2-E335G

The ESF1 crystal form proved to be the most practical of the three forms and hence it was this form that was selected for further crystallographic characterization of Xyn52B2-WT, Xyn52B2-E335G and Xyn52B2-E335G-Se. After refinement, the optimal crystallization conditions for all three proteins appeared to be for protein concentrations of 4.7 mg ml⁻¹, over a reservoir of 17–20% PEG 3K, 200 mM calcium acetate, 0.1 M Tris buffer pH 8.5. These crystals grew to their full size after about 5–10 d, with typical dimensions in the range 0.02–0.05 × 0.02–0.05 × 0.07–0.3 mm (Figs. 1*a* and 1*b*). Within these time and size ranges, the crystals obtained for Xyn52B2-E335G appeared earlier and were usually the largest, and those of Xyn52B2-E335G-Se appeared last and were usually the smallest. The crystallization times and sizes for Xyn52B2-WT were roughly in the middle of these ranges.

Several of these Xyn52B2-E335G ESF1 crystals were selected for detailed crystallographic characterization and measurement of X-ray diffraction data under cryogenic conditions. These experiments were initially performed at the TCSB (Technion, Haifa) using an FR-X rotating-anode (Rigaku, Japan) X-ray source ($\lambda = 1.54$ Å) coupled to an R-AXIS HTC (Rigaku, Japan) imaging-plate detector and an Oxford Cryosystems crystal-cooling system (100 K). The crystal-cooling procedure used for these experiments included a short soak of the target crystal (about 30–60 s) in a cryoprotecting solution consisting of the original crystallization mother liquor with 20% (v/v) additional PEG 200. Each of these pre-soaked crystals were then submitted to the flash-cooling procedure by immediate immersion in liquid nitrogen followed by quick transfer into the centre of a cold nitrogen-gas stream (100 K) flowing around the crystal throughout the X-ray data collection. The observed diffraction pattern for the ESF1 crystals at the TCSB was generally rather weak and usually did not show significant diffraction beyond 3.5 Å resolution. Nevertheless, some of these crystals diffracted to around 3.0 Å resolution and their diffraction pattern indicated that the crystals belonged to a primitive orthorhombic crystal system (space group $P2_12_12_1$), with average unit-cell parameters $a = 97.7$, $b = 119.1$, $c = 242.3$ Å. Different crystals of this crystal form gave similar unit-cell parameters; however, while the overall deviation from these average values for a and b was of less than 0.5%, the observed deviations for c could be as much as 1.5%.

One of these ESF1 crystals (Figs. 1*a* and 1*b*) was used for complete X-ray diffraction data collection at 2.95 Å resolution using the TCSB crystallographic setup (see above). The raw imaging-plate diffraction images were processed and integrated with the *HKL-3000* program suite (Minor *et al.*, 2006) and the representative data-collection parameters are summarized in Table 1. The relatively high R_{merge} (15.4%) probably originated from the rather weak diffraction pattern and the relatively high degree of anisotropy. While reasonable diffraction could be measured to 2.95 Å resolution for some of the crystal orientations (data not shown), considerably weaker diffraction was observed in other crystal orientations, especially in the outer diffraction shells. Moreover, significant radiation damage was noticed during the rather long data collection (about 22 h), a factor that could also lead to non-ideal values of R_{merge} and diffraction completeness.

At this point, it was possible to estimate the protein content in the crystallographic unit cell by a rough calculation of the solvent content in the crystal and the specific ratio of volume to protein mass (V_M). The range of V_M values observed originally for soluble protein

crystals was 1.68–3.5 Å³ Da⁻¹ with a mean of 2.17 Å³ Da⁻¹ (based on 116 proteins; Matthews, 1968). A wider range and a mean of 2.69 Å³ Da⁻¹ was subsequently calculated based on 10 471 proteins (Kantardjieff & Rupp, 2003). These mean V_M values correspond to mean solvent contents of 43 and 47%, respectively. The volume of the Xyn52B2-E335G unit cell (in the ESF1 crystal form), as determined from the mean value of the unit-cell parameters at 100 K, is 2.82×10^6 Å³. Assuming that the V_M value here is within the normal range, there should be between three and five Xyn52B2 monomers (705 amino acids; molecular weight 79 899) in the asymmetric unit. With three molecules in the ESF1 asymmetric unit (12 in the unit cell), the calculated V_M is 2.95 Å³ Da⁻¹ and the corresponding solvent content is 58.26%. With four molecules in the asymmetric unit (16 in the unit cell), the calculated V_M is 2.21 Å³ Da⁻¹ and the corresponding solvent content is 44.35%. With five molecules in the asymmetric unit (20 in the unit cell), the calculated V_M is 1.77 Å³ Da⁻¹ and the corresponding solvent content is 30.44%. These three different possibilities all seem theoretically possible considering their calculated V_M values and solvent contents; however, the second possibility (four monomers per asymmetric unit) gives values that are significantly closer to the mean values observed previously, making four monomers per asymmetric unit the most likely solution. Such a content of the asymmetric unit is also consistent with gel-filtration experiments, which suggest that Xyn52B2 is active as a dimer (see above), making two dimers in the asymmetric unit the most reasonable option.

3.3. X-ray diffraction data for Xyn52B2-WT

As mentioned above, the native Xyn52B2 protein (Xyn52B2-WT) crystallized under very similar conditions and with a similar crystal shape to the ESF1 crystal form of the Xyn52B2-E335G protein, although only after significantly longer crystallization times and with significantly smaller crystal dimensions. Several such ESF1-form crystals of Xyn52B2-WT have been submitted to detailed crystallographic characterization with a synchrotron-radiation source (BM14, ESRF, Grenoble, France) and a CCD area detector (MAR 225, MAR Research, USA). The crystal-cooling procedure was similar to that outlined above. These crystals were shown to be closely isomorphous to those of the E335G mutant described above, with an identical space group ($P2_12_12_1$) and very similar unit-cell parameters (average of $a = 97.5$, $b = 120.0$, $c = 238.3$ Å; Table 1). As expected from their smaller diffracting volumes, the observed diffraction pattern of these crystals was significantly weaker than that of the corresponding E335G crystals, with useful diffraction limits in the 4.0–3.7 Å resolution range. One of these Xyn52B2-WT crystals was used for complete X-ray diffraction data collection at 3.70 Å resolution on the BM14 beamline at the ESRF ($\lambda = 0.979$ Å, 100 K). Data processing was performed with *iMosflm* (Battye *et al.*, 2011) and the data-collection parameters are summarized in Table 1.

3.4. X-ray diffraction data for Xyn52B2-E335G-Hg

At the time of data collection, no protein structure with sufficient homology to Xyn52B2 was available to allow structure solution *via* molecular-replacement techniques. An attempt was therefore undertaken to prepare heavy-atom derivatives of the protein (wild type or mutants) in order to use them for phase determination *via* traditional isomorphous replacement procedures. One of these experiments involved the specially synthesized mercury-substituted xylose derivative described above [1-(thiomethylmercuric)- β -D-xyloside; X-S-Hg-CH₃], assuming that since the D-xylose monomer is one of the products of the usual xylosidase catalytic reaction of

Xyn52B2, a substituted D-xylose such as X-S-Hg-CH₃ will bind to its active site with reasonable affinity. These experiments were mainly carried out with the Xyn52B2-E335G mutant since it routinely produced significantly better crystals. In the particular case described here, we soaked a well formed crystal of Xyn52B2-E335G in a cryoprotecting solution which contained the X-S-Hg-CH₃ compound at 10 mM. The soak was performed for about 60 s prior to crystal cooling, using the same flash-cooling procedure described above. An

energy scan was performed on the mounted crystal, validating that mercury was present in the frozen sample (see reservations above), and the crystal was then submitted to detailed crystallographic characterization (BM14, ESRF, Grenoble, France). The crystal treated with this procedure proved to be completely isomorphous to the unsoaked ESF1-form Xyn52B2-E335G crystals, with very similar unit-cell parameters in space group $P2_12_12_1$ of $a = 97.4$, $b = 119.1$, $c = 243.0$ Å. A complete X-ray diffraction data collection was performed

Figure 2

X-ray diffraction pattern of Xyn52B2-E335G-Hg (ESF1 form) obtained using a synchrotron source (BM14, ESRF). The outer circle corresponds to 2.2 Å resolution. The insets represent magnified views of the sections indicated by the corresponding squares (bottom, low resolution; top, medium resolution).

Figure 3

X-ray diffraction pattern of Xyn52B2-E335G-Se (ESF1 form) obtained using a synchrotron source (BM14, ESRF). The outer circle corresponds to 3.9 Å resolution. The insets represent magnified views of the sections indicated by the corresponding squares (top, low resolution; bottom, medium resolution). It is noted that in principle the diffraction pattern exceeds 3.9 Å resolution, but the diffraction spots in this range are rather diffuse and were therefore not included in the final data set.

on this crystal to 2.15 Å resolution ($\lambda = 0.978$ Å; Fig. 2) and data processing was performed with *iMosflm* (Battye *et al.*, 2011), as summarized in Table 1.

3.5. X-ray diffraction data for Xyn52B2-E335G-Se

A selenium derivative of Xyn52B2 was considered to be another alternative for the structure determination of the protein, as the crystallographic phase problem could be resolved by either the single-wavelength or multi-wavelength anomalous diffraction (SAD or MAD) techniques. A fully replaced selenomethionine derivative of Xyn52B2-E335G (Xyn52B2-E335G-Se) was therefore over-expressed and purified as described above, resulting in a protein with 16 Se atoms per protein monomer. Crystals of this selenomethionine-derivatized protein were obtained using the same general conditions that produced the ESF1-form crystals of the Xyn52B2-E335G protein, although with much longer crystallization times and with much smaller final crystals. The relatively weak diffraction patterns of these crystals, some of which exceeded 4.0 Å resolution, confirmed that they are highly isomorphous to the corresponding ESF1-form crystals of Xyn52B2-E335G (see above), as reflected by the same space group ($P2_12_12_1$) and very similar unit-cell parameters (Table 1). One of these Xyn52B2-E335G-Se crystals was used to collect a complete SAD data set at the selenium absorption peak ($\lambda = 0.978$ Å, 3.90 Å resolution; Fig. 3) (BM14, ESRF, Grenoble, France). Data processing was performed with the *HKL-2000* suite (Otwinowski & Minor, 1997), and the corresponding data-collection and processing parameters are summarized in Table 1. The Xyn52B2 monomer contains 16 methionines, providing 16 Se atoms in the Xyn52B2-Se derivative per protein monomer of 705 amino-acid residues. As a general rule of thumb, for a successful MAD or SAD experiment on a selenomethionine-containing protein, at least one selenomethionine per 100 amino acids is required (Hendrickson & Ogata, 1997; Dauter, 2006). Thus, this Xyn52B2-Se derivative has a Se:protein ratio that should be sufficient for reasonable phasing *via* the Se-SAD methodology.

Whether provided by the Hg-derivative or the Se-SAD data, the resulting crystallographic phasing should lead to structure determination of the corresponding Xyn52B2-E335G-Hg or Xyn52B2-E335G-Se proteins at 2.15 or 3.90 Å resolution, respectively. Once obtained, these models could lead to structural analysis of the wild-type protein at 3.70 Å resolution and of the E335G mutant at 2.95 Å resolution. Recently, however, the structure of a highly homologous protein structure has been determined, a β -xylosidase from *G. thermoglucosidasius* that shares 86% sequence homology with Xyn52B2 (PDB entry 4c1o; Espina *et al.*, 2014) providing an important reference structure that was not available at the time of our original diffraction data measurement. This structure should be a good model for molecular-replacement procedures, which would probably be the most efficient method to determine the structure of Xyn52B2 with the diffraction data described for Xyn52B2-WT and its mutant and derivatives. Such analyses are currently in progress in our laboratory.

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 of the TCSB (Technion, Haifa, Israel) for helpful assistance in the crystallization screening experiments. We thank the staff of the BM14 beamline at the ESRF and the corresponding EMBL/ESRF Outstation for their helpful support during the X-ray synchrotron data collection and analysis. The synchrotron experiments at the ESRF were also supported by the ESRF internal funding program. YS holds the Erwin and Rosl Pollak Chair in Biotechnology at the Technion.

References

- Alalouf, O., Balazs, Y., Volkinshtein, M., Grimpel, Y., Shoham, G. & Shoham, Y. (2011). *J. Biol. Chem.* **286**, 41993–42001.
- Almog, O., Greenblatt, H. M., Spungin, A., Ben-Meir, D., Blumberg, S. & Shoham, G. (1993). *J. Mol. Biol.* **230**, 342–344.
- Almog, O., Klein, D., Braun, S. & Shoham, G. (1994). *J. Mol. Biol.* **235**, 760–762.
- Bar, M., Golan, G., Nechama, M., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2004). *Acta Cryst.* **D60**, 545–549.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Belakhov, V., Dor, E., Hershenhorn, J., Botoshansky, M., Bravman, T., Kolog, M., Shoham, Y., Shoham, G. & Baasov, T. (2001). *Isr. J. Chem.* **40**, 177–188.
- 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.
- 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.
- Brux, C., Ben-David, A., Shallom-Shezifi, D., Leon, M., Niefind, K., Shoham, G., Shoham, Y. & Schomburg, D. (2006). *J. Mol. Biol.* **359**, 97–109.
- Czjzek, M., Bravman, T., Henrissat, B. & Shoham, Y. (2004). *Acta Cryst.* **D60**, 1461–1463.
- Dauter, Z. (2006). *Acta Cryst.* **D62**, 867–876.
- Donk, P. J. (1920). *J. Bacteriol.* **5**, 373–374.
- Espina, G., Eley, K., Pompidor, G., Schneider, T. R., Crennell, S. J. & Danson, M. J. (2014). *Acta Cryst.* **D70**, 1366–1374.
- Gat, O., Lapidot, A., Alchanati, I., Regueros, C. & Shoham, Y. (1994). *Appl. Environ. Microbiol.* **60**, 1889–1896.
- Gilboa, R., Bauer, A. J. & Shoham, G. (1998). *Acta Cryst.* **D54**, 1467–1470.
- 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.
- Golan, G., Zharkov, D. O., Fernandes, A. S., Zaika, E., Kycia, J. H., Wawrzak, Z., Grollman, A. P. & Shoham, G. (2004). *Acta Cryst.* **D60**, 1476–1480.
- Hancock, S. M., Vaughan, M. D. & Withers, S. G. (2006). *Curr. Opin. Chem. Biol.* **10**, 509–519.
- Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
- Hövel, K., Shallom, D., Niefind, K., Baasov, T., Shoham, G., Shoham, Y. & Schomburg, D. (2003). *Acta Cryst.* **D59**, 913–915.
- Hövel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. & Schomburg, D. (2003). *EMBO J.* **22**, 4922–4932.
- Huang, Z., Liu, X., Zhang, S. & Liu, Z. (2014). *J. Ind. Microbiol. Biotechnol.* **41**, 65–74.
- Jordan, D. B., Vermillion, K. E., Grigorescu, A. A. & Braker, J. D. (2013). *Arch. Biochem. Biophys.* **537**, 176–184.
- Kantardjiev, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Lansky, S., Alalouf, O., Salama, R., Dvir, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* **F70**, 476–481.
- Lansky, S., Alalouf, O., Solomon, V., Alhassid, A., Govada, L., Chayen, N. E., Belrhali, H., Shoham, Y. & Shoham, G. (2013). *Acta Cryst.* **F69**, 430–434.
- Lansky, S., Alalouf, O., Solomon, V. H., Alhassid, A., Govada, L., Chayen, N. E., Belrhali, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* **D70**, 261–278.
- Lansky, S., Salama, R., Dann, R., Shner, I., Manjasetty, B. A., Belrhali, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* **F70**, 1038–1045.
- Lansky, S., Salama, R., Solomon, V. H., Belrhali, H., Shoham, Y. & Shoham, G. (2013). *Acta Cryst.* **F69**, 695–699.
- Lansky, S., Zehavi, A., Dann, R., Dvir, H., Belrhali, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* **F70**, 225–231.

- Lundgren, K. R., Bergkvist, L., Hogman, S., Joves, H., Eriksson, G., Bartfai, T., van der Laan, J., Rosenberg, E. & Shoham, Y. (1994). *FEMS Microbiol. Rev.* **13**, 365–368.
- Mackenzie, L. F., Wang, Q., Warren, R. A. J. & Withers, S. G. (1998). *J. Am. Chem. Soc.* **120**, 5583–5584.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Mechaly, A., Teplitsky, A., Belakhov, V., Baasov, T., Shoham, G. & Shoham, Y. (2000). *J. Biotechnol.* **78**, 83–86.
- Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* **D62**, 859–866.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Rees, D. C., Johnson, E. & Lewinson, O. (2009). *Nature Rev. Mol. Cell Biol.* **10**, 218–227.
- Reiland, V., Fundoiano-Herscovitz, Y., Golan, G., Gilboa, R., Shoham, Y. & Shoham, G. (2004). *Acta Cryst.* **D60**, 2371–2376.
- 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.
- Shoham, Y., Schwartz, Z., Khasin, A., Gat, O., Zosim, Z. & Rosenberg, E. (1992). *Biodegradation*, **3**, 207–218.
- 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., Shenker, O., Langut, Y., Lavid, N., Gat, O., Zaide, G., Zehavi, A., Sonenshein, A. L. & Shoham, Y. (2014). *J. Biol. Chem.* **289**, 25957–25975.
- Shulami, S., Zaide, G., Zolotnitsky, G., Langut, Y., Feld, G., Sonenshein, A. L. & Shoham, Y. (2007). *Appl. Environ. Microbiol.* **73**, 874–884.
- Solomon, H. V., Tabachnikov, O., Feinberg, H., Govada, L., Chayen, N. E., Shoham, Y. & Shoham, G. (2013). *Acta Cryst.* **F69**, 1114–1119.
- Solomon, V., Teplitsky, A., Shulami, S., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2007). *Acta Cryst.* **D63**, 845–859.
- Suurnäkki, A., Tenkanen, M., Buchert, J. & Viikari, L. (1997). *Adv. Biochem. Eng. Biotechnol.* **57**, 261–287.
- 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.* **D53**, 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.* **D60**, 836–848.
- Teplitsky, A., Shulami, S., Moryles, S., Shoham, Y. & Shoham, G. (2000). *Acta Cryst.* **D56**, 181–184.
- Teplitsky, A., Shulami, S., Moryles, S., Zaide, G., Shoham, Y. & Shoham, G. (1999). *Acta Cryst.* **D55**, 869–872.
- Yeoman, C. J., Han, Y., Dodd, D., Schroeder, C. M., Mackie, R. I. & Cann, I. K. (2010). *Adv. Appl. Microbiol.* **70**, 1–55.
- Zaide, G., Shallom, D., Shulami, S., Zolotnitsky, G., Golan, G., Baasov, T., Shoham, G. & Shoham, Y. (2001). *Eur. J. Biochem.* **268**, 3006–3016.