Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

A unique octameric structure of Axe2, an intracellular acetyl-xylooligosaccharide esterase from *Geobacillus stearothermophilus*

Shifra Lansky,[a] Onit Alalouf,[b] Hodaya Vered Solomon,[a] Anat Alhassid,[a] Lata Govada,[c] Naomi E. Chayen,[c] Hassan Belrhali,[d] Yuval Shoham[b]* and Gil Shoham[a]*

[a]Institute of Chemistry and the Laboratory for Structural Chemistry and Biology, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel, [b]Department of Biotechnology and Food Engineering, Technion – Israel Institute of Technology, 32000 Haifa, Israel, [c]Department of Surgery and Cancer, Imperial College London, London SW7 2AZ, England, and [d]European Synchrotron Radiation Facility, BP 220, 38043 Grenoble, France

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

Received 13 September 2013
Accepted 15 October 2013

PDB references: Axe2-WT, 3w7v; Axe2-Se, 4jhl; Axe2-S15A, 4jko

Geobacillus stearothermophilus T6 is a thermophilic, Gram-positive soil bacterium that possesses an extensive and highly regulated hemicellulolytic system, allowing the bacterium to efficiently degrade high-molecular-weight polysaccharides such as xylan, arabinan and galactan. As part of the xylan-degradation system, the bacterium uses a number of side-chain-cleaving enzymes, one of which is Axe2, a 219-amino-acid intracellular serine acetylxylan esterase that removes acetyl side groups from xylooligosaccharides. Bioinformatic analyses suggest that Axe2 belongs to the lipase GDSL family and represents a new family of carbohydrate esterases. In the current study, the detailed three-dimensional structure of Axe2 is reported, as determined by X-ray crystallography. The structure of the selenomethionine derivative Axe2-Se was initially determined by single-wavelength anomalous diffraction techniques at 1.70 Å resolution and was used for the structure determination of wild-type Axe2 (Axe2-WT) and the catalytic mutant Axe2-S15A at 1.85 and 1.90 Å resolution, respectively. These structures demonstrate that the three-dimensional structure of the Axe2 monomer generally corresponds to the SGNH hydrolase fold, consisting of five central parallel β-sheets flanked by two layers of helices (eight α-helices and five 3_{10}-helices). The catalytic triad residues, Ser15, His194 and Asp191, are lined up along a substrate channel situated on the concave surface of the monomer. Interestingly, the Axe2 monomers are assembled as a 'doughnut-shaped' homo-octamer, presenting a unique quaternary structure built of two staggered tetrameric rings. The eight active sites are organized in four closely situated pairs, which face the relatively wide internal cavity. The biological relevance of this octameric structure is supported by independent results obtained from gel-filtration, TEM and SAXS experiments. These data and their comparison to the structural data of related hydrolases are used for a more general discussion focusing on the structure–function relationships of enzymes of this category.

1. Introduction

Acetylxylan esterases are enzymes that remove acetyl side chains from the backbone of xylan, the main hemicellulose in hardwoods and annual plants. Xylan consists of a linear backbone of β-(1,4)-D-xylopyranosyl residues, which can be decorated by α-L-arabinofuranosyl units in the O-2 and/or O-3 positions, by 4-*O*-methylglucopyranosyl uronic acid in the O-2 position and/or by acetyl groups in the O-2 and/or O-3 positions (Aspinall, 1959). In hardwood glucuronoxylans, about 70% of the xylopyranosyl residues are acetylated at the C-2 or C-3 position or both (Bouveng, 1961; Lindberg *et al.*, 1973*a*), whereas in softwood glucomannans the acetyl content is lower (Lindberg *et al.*, 1973*b*). The removal of the acetyl groups in

xylan has been shown to improve the ability of the main-chain hydrolyzing enzymes to break down the polymer (Biely, 2012; Zhang *et al.*, 2011). A part of the acetylxylan esterases studied and reported to date have been classified in the CAZy database and are found in eight of the 16 carbohydrate esterase (CE) families characterized to date (Cantarel *et al.*, 2009). Families CE3 (acetylxylan esterases) and CE12 (pectin acetylesterases, rhamnogalacturonan acetylesterases and acetylxylan esterases) are also classified as lipase GDSL family proteins (Pfam accession No. PF00657).

Geobacillus stearothermophilus T6 is a thermophilic soil bacterium that possesses an extensive and highly regulated hemicellulolytic system that allows the bacterium to efficiently compete for the scarce carbon in the soil (Shulami *et al.*, 1999, 2007; Tabachnikov & Shoham, 2013). Under the appropriate conditions, the bacterium can secrete up to three extracellular enzymes that hydrolyze the high-molecular-weight polysaccharides xylan, arabinan and galactan (Gat *et al.*, 1994; Khasin *et al.*, 1993; Lapidot *et al.*, 1996; Bar *et al.*, 2004). For the utilization of xylan, the bacterium secretes an extracellular GH10 xylanase, XT6 (Gat *et al.*, 1994; Khasin *et al.*, 1993; Lapidot *et al.*, 1996; Bar *et al.*, 2004; Zolotnitsky *et al.*, 2004), which degrades the polysaccharide to decorated xylooligomers that are transported into the cell *via* a specific ATP-binding cassette (ABC) sugar-transport system (Shulami *et al.*, 2007). Once inside the cell, the decorated xylooligomers are hydrolyzed by side-chain-cleaving enzymes, including two GH51 α-arabinofuranosidases (Shallom, Belakhov, Solomon, Gilead-Gropper *et al.*, 2002; Shallom, Belakhov, Solomon, Shoham *et al.*, 2002; Hövel *et al.*, 2003), a GH67 α-glucuronidase (Zaide *et al.*, 2001; Golan *et al.*, 2004; Shallom *et al.*, 2004) and two xylanesterases (Alalouf *et al.*, 2011), and finally by a GH10 intracellular xylanase, IXT6 (Solomon *et al.*, 2007), and three β-xylosidases (GH39, GH43 and GH52; Bravman, Mechaly *et al.*, 2001; Bravman, Zolotnitsky *et al.*, 2001; Bravman *et al.*, 2003; Czjzek *et al.*, 2004; Ben-David *et al.*, 2007). In addition, the bacterium utilizes two- or three-component oligosaccharide-sensing systems that activate the expression of the sugar transporters, allowing it to respond rapidly to very low concentrations of carbohydrates in the soil and to efficiently utilize them. This overall strategy has the advantage that the bacterium can utilize extracellular soluble carbohydrate matter that is not readily available to competing microorganisms.

We have recently identified the *axe2* gene (GenBank accession No. ABI49953.1) as part of a three-gene operon in *G. stearothermophilus*, which also includes *xynB3* (encoding a GH43 β-xylosidase; Brüx *et al.*, 2006; Shallom *et al.*, 2005) and an uncharacterized gene *xynX* (GenBank accession No. DQ868502.2). According to sequence similarities, the *axe2* gene product, Axe2 (UniProtKB/TrEMBL accession No. Q09LX1), is an intracellular serine hydrolase belonging to the lipase GDSL-2 family (PF13472). The monomer of this protein is made up of 219 amino acids with a calculated molecular mass of 24 770 Da. The lipase GDSL-2 family is one of eight families that make up the SGNH hydrolase superfamily (Pfam clan accession No. CL0264; Akoh *et al.*, 2004; Finn *et al.*, 2010). Members of the SGNH superfamily possess four strictly conserved residues: Ser, Gly, Asn and His in four conserved blocks I, II, III and V, respectively (Akoh *et al.*, 2004). Each of the four residues plays a key role in the catalytic function of the enzyme (Ser as the nucleophile, Gly and Asn as proton donors to the oxyanion hole, and His as the acid/base in the catalytic triad). Members of the GDSL family possess the GDS*X* motif, which is equivalent to the classical G*X*S*X*G motif of lipases/esterases, in which the Ser of the GDS*X* motif is the nucleophilic Ser in block I of the protein.

Axe2 was recently characterized biochemically in our laboratory (Alalouf *et al.*, 2011), demonstrating that the purified protein completely deacetylates xylobiose peracetate (fully acetylated xylobiose) and hydrolyzes the synthetic substrates 2-naphthyl acetate, 4-nitrophenyl acetate, 4-methylumbelliferyl acetate and phenyl acetate. Based on the genetic context of the Axe2 gene and its experimental substrate specificity, Axe2 is determined to be a xylooligosaccharide esterase. Bioinformatics analyses suggest that Axe2 and its homologues do not belong to any known family in the CAZy database and thus represent a new family of carbohydrate esterases (Alalouf *et al.*, 2011).

In the framework of a larger research project aimed at studying the structure–function relationships of all of the xylan-degrading enzymes of *G. stearothermophilus*, we recently obtained high-quality single crystals of native Axe2 and several of its catalytic mutants, which enabled X-ray crystallographic diffraction measurements at high resolution (Lansky *et al.*, 2013). In the present report, we describe the crystallographic analysis and the resulting three-dimensional structure of the selenomethionine derivative Axe2-Se at 1.70 Å resolution, the structure of the native protein (Axe2-WT) at 1.85 Å resolution and the structure of the Axe2-S15A catalytic mutant at 1.90 Å resolution. These structures demonstrate that Axe2 is organized and probably functions as a 'doughnut-shaped' homo-octamer, presenting a unique quaternary structure built of two staggered tetrameric rings. This structure is used for comparison with related carbohydrate hydrolases in order to draw more general conclusions about the structure–specificity relationships in enzymes of this category.

2. Experimental methods

2.1. Crystal structure determination of Axe2-Se and Axe2-WT

The Axe2 protein represents a new carbohydrate esterase family (Alalouf *et al.*, 2011) and, at the time of the structural analysis, had no reference protein with sufficiently homologous structural features to allow solution by molecular replacement. Thus, the selenium single-wavelength anomalous diffraction (SAD) method was used instead in this case, as Axe2 contains seven methionine residues which can easily be replaced by selenomethionine residues. The expression, purification, crystallization and preliminary crystallographic analysis of native Axe2 (Axe2-WT) and the fully substituted

Table 1
Representative diffraction data-collection parameters for Axe2.

Values in parentheses are for the last resolution shell.

Protein	Axe2-WT	Axe2-S15A	Axe2-Se (edge)	Axe2-Se (remote)
Wavelength (Å)	0.954	0.954	0.978	0.954
Space group	*I*422	*I*422	*I*422	*I*422
Unit-cell parameters (Å)				
$a = b$	109.83	109.94	110.42	110.30
c	213.31	213.02	212.87	213.09
Resolution range (Å)	30.00–1.85 (1.88–1.85)	35.00–1.90 (1.93–1.90)	30.0–1.85 (1.88–1.85)	30.0–1.70 (1.73–1.70)
Multiplicity	9.8 (8.3)	8.1 (7.7)	29.5 (29.3)	8.2 (8.1)
$\langle I\rangle/\langle\sigma(I)\rangle$	10.8 (3.3)	10.8 (3.9)	6.4 (5.2)	8.3 (3.5)
Mosaicity (°)	0.310	0.306	0.243	0.318
Completeness (%)	100 (100)	100 (100)	100 (100)	98.9 (99.9)
Square $R_{\rm sym}$ (%)†	5.8 (47.5)	6.6 (40.2)	7.6 (51.2)	5.6 (46.1)

† Square $R_{\rm sym} = \sum(I - \langle I\rangle)^2/\sum I^2$.

selenomethionine derivative of Axe2 (Axe2-Se) were generally carried out as described previously (Lansky *et al.*, 2013). The relevant diffraction data-collection and processing details are presented in Table 1.

The structure of Axe2 was determined by SAD phasing using Se edge and remote diffraction data sets obtained from the Axe2-Se crystals (1.85 and 1.70 Å resolution, respectively; Lansky *et al.*, 2013). Se-atom sites, density modification and initial phases were produced using the *SHELX* suite of software embedded within the *HKL2MAP* v.0.3.a-beta GUI (Pape & Schneider, 2004). Using *SHELXC* (Sheldrick, 2008), very strong anomalous signal to high resolution was detectable (d''/σ varied from 7.3 to 2.0 at 20 to 1.85 Å resolution, respectively). The Matthews coefficient calculations suggested that two molecules compose the asymmetric unit (solvent content of 55%). Consequently, 14 Se atoms were searched for and located using *SHELXD* (Sheldrick, 2008). A trivial solution with very high correlation coefficients (CC_{weak} = 38 and CC_{all} = 60) was determined for exactly the expected 14 selenium sites. Finally, *SHELXE* (Sheldrick, 2008, 2010) was used to calculate an initial set of phases based on the 14 selenium positions (both hands) and to carry out 20 cycles of density modification followed by three cycles of auto-tracing and model building. For this last step, the remote data set at 1.70 Å resolution was used and phase extension was performed up to this resolution. This procedure led to the building of an initial model of the Axe2 dimer, consisting of 420 polyalanine stretches (Supplementary Fig. S1[1]).

This initial model of the dimer was used for the full protein three-dimensional model-building step (*Coot*; Emsley *et al.*, 2010), based on the known amino-acid sequence of the Axe2 protein (Alalouf *et al.*, 2011) and the clear positions of the 14 selenomethionine residues. This initial model, consisting of two independent molecules in the asymmetric unit, was subjected to rigid-body refinement, simulated annealing and interactive refinement cycles of positional parameters and temperature factors with the crystallographic refinement

Table 2
Representative structure-determination and refinement parameters.

	Axe2-WT	Axe2-S15A	Axe2-Se
Model refinement			
Data resolution range (Å)	30.00–1.85	35.00–1.90	30.0–1.70
Data cutoff	$\|F\| > 0$	$\|F\| > 0$	$\|F\| > 0$
R factor† (%)	14.07	14.87	13.13
No. of reflections	53030	49014	67992
$R_{\rm free}$† (%)	17.72	18.84	16.39
No. of reflections	2836	2633	3655
Refined model			
No. of residues	438	438	438
No. of protein atoms	4122	4111	4219
No. of solvent atoms	521	523	548
Solvent content (%)	55.52	55.66	54.81
No. of molecules per asymmetric unit	2	2	2
Average B factor (Å^2)	29.0	32.5	25.7
R.m.s. deviations			
Bond lengths (Å)	0.022	0.022	0.024
Bond angles (°)	2.14	2.14	2.36
Estimated coordinate error, Luzzati (Å)	0.17	0.19	0.15
Ramachandran plot, % residues in			
Favoured region	96.5	97.5	97.9
Allowed region	3.5	2.5	2.1
Outlier region	0.0	0.0	0.0
PDB code	3w7v	4jko	4jhl

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

program *REFMAC*5 (Murshudov *et al.*, 2011). Further refinement and addition of water molecules were performed with the tandem programs *REFMAC*5 (Murshudov *et al.*, 2011) and *ARP*/*wARP* (Lamzin & Wilson, 1993), as available within the *CCP*4 suite (Winn *et al.*, 2001). Each round of refinement was followed by manual fitting and rebuilding of the protein model into the electron-density map with *Coot* (Emsley *et al.*, 2010). This procedure greatly improved the

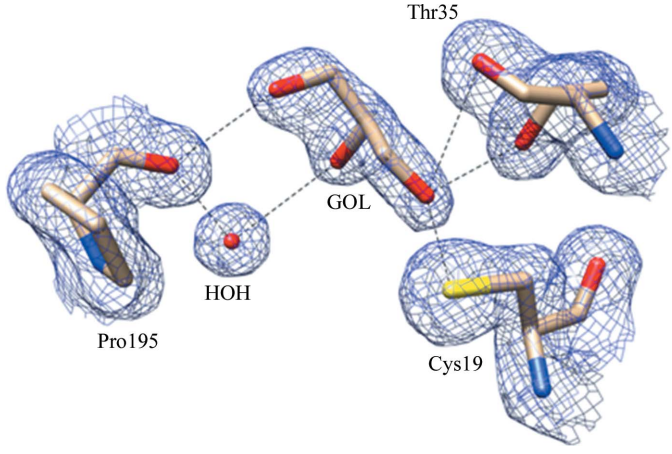


Figure 1
A representative section of the final electron-density map of Axe2-Se, demonstrating its high quality and the high reliability of the map interpretation. Superimposed on the map are the corresponding regions of the final model of Axe2 (conventional atom colours), showing residues Cys19, Thr35 and Pro195 bound to a glycerol molecule (GOL) and a water molecule (HOH). Hydrogen bonds are marked by dotted lines. The map corresponds to 1.70 Å resolution and is contoured at the 1.3σ level.

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: DW5078).

quality of the electron-density maps and allowed the reconstruction of all 438 (2 × 219) protein residues in the dimer. Water molecules were assigned to peaks in the ($F_o - F_c$) difference electron-density maps at contour levels greater than 3.4σ, which were also at a suitable distance and orientation to form a hydrogen bond to a potential partner. In the final cycle, 548 such water molecules were included in the dimeric crystallographic model of Axe2-Se. A representative section of the final electron-density map is shown in Fig. 1 and representative parameters for the refinement and the final model are listed in Table 2.

The fully refined three-dimensional structure of the Axe2-Se derivative at 1.70 Å resolution was used as the basis for structure determination of the native enzyme (Axe2-WT) at 1.85 Å resolution. The structure-determination and refinement procedures were generally similar to those used for Axe2-Se, except that the initial model started from a significantly better *R* factor of 22.03%. Representative parameters for the refinement are listed in Table 2.

2.2. Content of the final crystallographic model of Axe2-WT

As originally estimated, the asymmetric unit contains two independent molecules of the Axe2-WT monomer (Lansky *et al.*, 2013), labelled here as chain *A* and chain *B*. All amino-acid residues of the two chains (438 residues) are present in the final model of the protein, labelled here accordingly as Met1.*A* to Arg219.*A* for chain *A* and as Met1.*B* to Arg219.*B* for chain *B*. Structural superposition of these two independent chains demonstrates an almost perfect overlap, with an r.m.s.d. of 0.249 Å based on 219 C^{α} atoms, an r.m.s.d. of 0.293 Å based on all main-chain atoms and an r.m.s.d. of 0.720 Å based on all atoms. Thus, despite some fluctuations and conformational differences in the side chains, especially on the surface, the two protein monomers are practically identical in their three-dimensional structure.

The conformational flexibility of some of the side chains is further demonstrated by their relatively weak and disordered electron density in the crystallographic maps. The disorder observed around Ser218 and Arg219 may originate from their location at the end of the polypeptide chain, while the disorder observed around Gly28, Ser29 and Phe30 may indicate a movement of the loop on which they are situated, as discussed below. Seven side chains in molecule *A* and eight side chains in molecule *B* were modelled in two alternate conformations. A *cis*-peptide bond was found between residues Arg23 and Pro24.

521 water molecules were unequivocally identified in the asymmetric unit (per protein dimer) and refined in the final model of Axe2-WT. Additional significant electron densities in the crystallographic $F_o - F_c$ maps were assigned as two chloride ions (in the active site) and five glycerol molecules (around the surface). These assignments were based on the magnitude and shape of the difference electron density, contacts with neighbouring functional groups, typical hydrogen-bond distances and typical halogen–ligand distances. The chloride ions found in the final model probably originated from the NaCl present in the crystallization solution of Axe2-WT, while the glycerol molecules most likely originated from the cryoprotecting solution (which contained 20% glycerol) in which the crystal was soaked prior to the data measurements (Lansky *et al.*, 2013).

The total content of the final crystallographic asymmetric unit of the Axe2-WT structure, as refined in the present work, includes 4122 (non-H) protein atoms (from the two protein monomers) and 553 (non-H) non-protein atoms, resulting in a total of 4675 non-H atoms per asymmetric unit.

2.3. Quality of the final model of Axe2-WT

The high quality of the current structure of Axe2-WT is demonstrated by its clear and easily interpretable electron-density maps (Fig. 1). The program *PROCHECK* (Laskowski *et al.*, 1993) was used for the validation of all structural parameters and stereochemical calculations. The Ramachandran plot (Ramachandran *et al.*, 1963) for Axe2-WT shows that 96.5% of the residues are in the most favoured regions, 3.5% are in the additionally allowed regions and no residues are in disallowed regions, a quite normal distribution of main-chain conformational parameters for medium-sized proteins in this resolution range. The protein molecules conform closely to standard bond lengths and angles, as defined by Engh & Huber (1991), with r.m.s.d.s of 0.022 Å and 2.14°, respectively. The overall average *B* factor is 29.0 Å^2. Based on the resolution of Axe2-WT (1.85 Å) and the final *R* factor (14.07%), the average experimental error in the coordinates of the final model is about 0.17 (±0.07) Å, according to the Luzzati error estimation (Luzzati, 1952), permitting a meaningful and reliable analysis of the interactions and geometries involved in the detailed three-dimensional structure presented here.

2.4. The Axe2-S15A catalytic mutant

Since Ser15 was identified as the critical nucleophile of the catalytic triad of the enzyme (Alalouf *et al.*, 2011), the first catalytic mutant produced for biochemical and structural studies was that of the S15A point mutation. The Axe2-S15A catalytic mutant was expressed and purified in a similar manner to Axe2-WT. Optimal crystals of Axe2-S15A were obtained using a slightly different crystallization procedure based on starting condition No. 2 (0.25 *M* NaH_2PO_4/1.0 *M* K_2HPO_4, 0.1 *M* imidazole buffer pH 8.0; Lansky *et al.*, 2013). Despite the different crystallization conditions, the resulting crystals of the Axe2-S15A mutant were confirmed to be highly isomorphous to the WT enzyme. Data collection, structure determination and refinement were performed according to the general procedure described above for Axe2-WT, and the corresponding parameters are summarized in Tables 1 and 2.

The crystal structure of the Axe2-S15A mutant does not deviate significantly from the structure of the native protein, except for the obvious differences in the mutation site and some conformational shifts in the flexible loop carrying residues Gly28, Ser29 and Phe30 (see below). These similarities further confirm the validity, stability and biological

significance of the Axe2 crystal structure, especially since the Axe2-S15A mutant was crystallized using significantly different experimental conditions.

Small local differences are observed in the two anion-binding sites discussed above. While in Axe2-WT and Axe2-Se these anions are identified as chloride ions, in the case of Axe2-S15A they are identified as phosphate ions (see below). These differences probably originate from the different crystallization conditions of the proteins, combined with the wider space available at the active site of the mutant owing to the shorter side chain at the mutation point (Ala *versus* Ser). In addition, a glycerol molecule was modelled near the loop formed by residues Gly28, Ser29 and Phe30 in Axe2-S15A, while in Axe2-Se a chloride ion was found instead. No non-protein molecules were found in this position in the case of Axe2-WT. In the Axe2-S15A model, an additional phosphate molecule was found coordinated to Arg149 NH1 and NH2 and Gln153 OE1. In total, 5–7 glycerol molecules are present in the three Axe2 structures reported here, 2–4 chloride ions are present in the Axe2-WT and Axe2-Se structures and three phosphate ions are present in the Axe2-S15A structure. The possible significance of these bound ions and glycerol molecules is discussed below.

2.5. Gel filtration

The molecular weight of Axe2-WT in solution was determined by gel filtration using an ÄKTA Explorer system equipped with a Superose 12 gel-filtration column (GE Healthcare Life Sciences; 24 ml total column volume). Protein samples (200 µl) were applied onto the column and eluted at room temperature (0.5 ml min^{-1}) with 50 m*M* Tris–HCl buffer pH 7 containing 0.1 *M* NaCl and 0.02% sodium azide. Molecular weights were determined from regression analysis of the log relative molecular weight 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*) were the extracellular xylanase XT6 (43.8 kDa), the extracellular β-1,4-galactanase GanA (87 kDa), the intracellular β-galactosidase GanB (240 kDa) and the intracellular xylosidase XynB2 (160 kDa) (Supplementary Fig. S2).

2.6. Transmission electron microscopy (TEM)

Negative-staining transmission electron microscopy was carried out at the Electron Microscopy Center of Soft Matter,

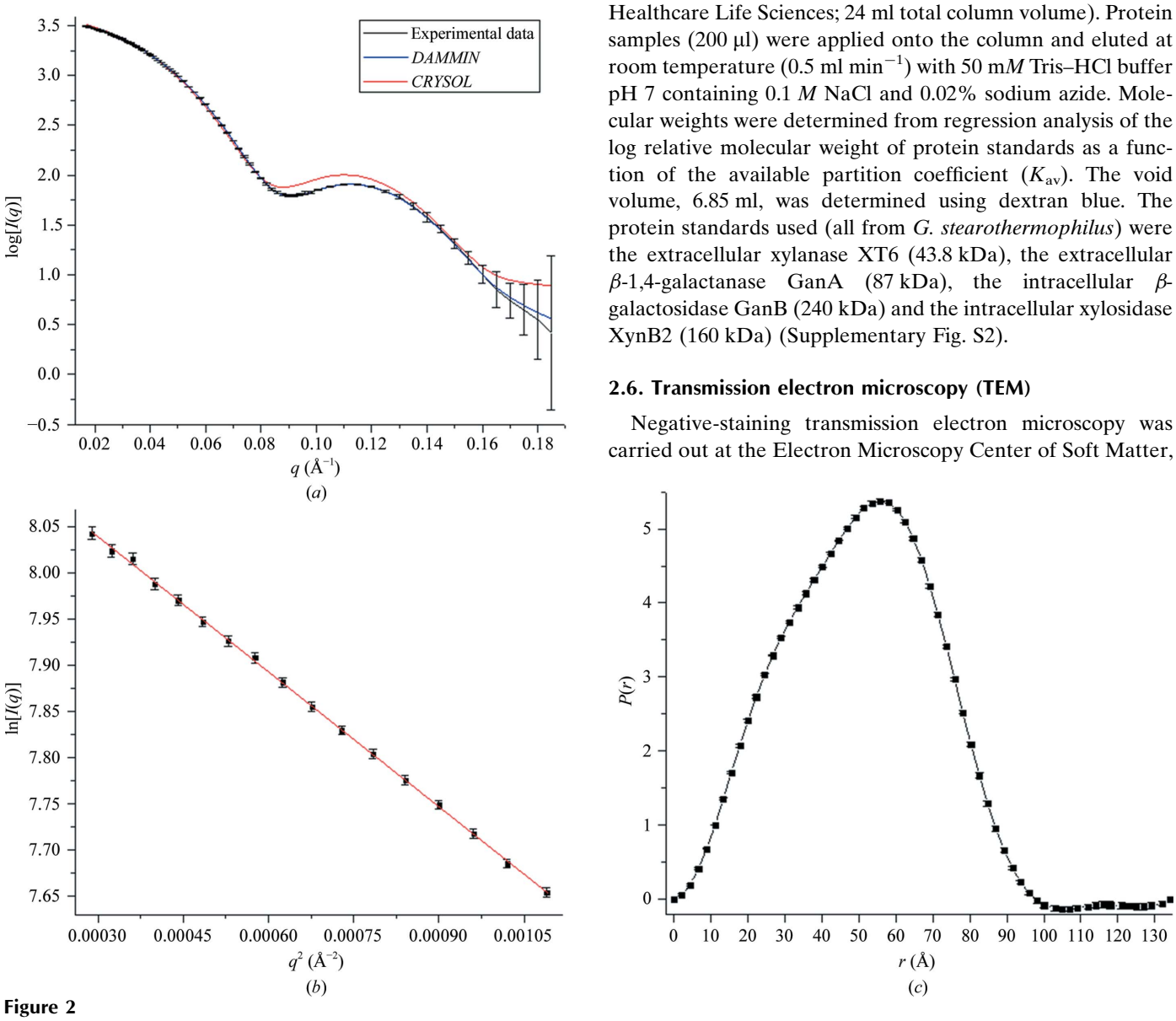


Figure 2
SAXS data for Axe2-WT. (*a*) Scattering curve for Axe2-WT. Log[$I(q)$] is plotted as a function of the momentum-transfer vector q. The q range taken for data analysis was q = 0.01–0.185 Å^{-1}. Experimental data are shown in black. The blue curve corresponds to the simulated scattering curve for one of the molecular-envelope models constructed using *DAMMIN* with χ^2 = 0.9192. The red curve corresponds to the simulated scattering curve from the crystallographic octameric model of Axe2 presented above and calculated by *CRYSOL* with χ = 5.86. (*b*) Guinier plot for Axe2-WT, showing a good linear fit (adjusted R^2 = 0.9996) within the range qR_g = 0.65–1.26, indicating no aggregation of the protein sample. (*c*) Pair-distance distribution function $P(r)$ for Axe2-WT.

Technion. 200-mesh grids, coated with carbon type B, were cleaned in a glow-discharge cleaning system (PELCO easi-Glow 91000, Ted Pella Inc., California, USA) just before use to increase their hydrophilicity. Samples were prepared as duplicates. A grid was incubated on the surface of a 20 µl drop of sample placed on Parafilm with the coated side facing the drop. After 1 min the grid was drawn from the sample drop, the excess sample was absorbed with a filter paper and the grid was incubated on a 20 µl drop of 2% uranyl acetate stain for 2 min and then transferred to another 20 µl drop of the stain for an additional 2 min incubation. The grid was then held with fine tweezers, clamped with a rubber ring, the stain was absorbed with the filter paper and the sample was dried at room temperature. Samples were examined using an FEI Tecnai G2 120 kV transmission electron microscope operating at 120 kV. Images were recorded using a Gatan US1000 high-resolution cooled CCD camera and were processed with *DigitalMicrograph* v.3.3.1.

2.7. Axe2-WT SAXS data

2.7.1. SAXS data collection. Small-angle X-ray scattering (SAXS) data were collected for Axe2-WT on beamline X9 of 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, 100 m*M* NaCl, 0.02% sodium azide at a protein concentration of 6.5 mg ml^{-1}. For buffer scattering subtractions, identical buffer samples were prepared and their scattering was measured immediately before measurement of the corresponding protein samples. Each measurement consisted of 30 s X-ray exposure (400 × 200 µm beam; $\lambda = 0.918$ Å) of a 20 µl sample flowing continuously through a 1 mm diameter capillary. This was repeated three times for accuracy. The measured scattered intensity ranged from $q = 0.01$ to 1.8 Å^{-1} for both proteins. The small-angle scattering intensity, $I(q)$, was measured for scattering angles of $q = 0.01$–1.8 Å^{-1} 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, 2010) was used for the wide-angle part of the scattering (Allaire & Yang, 2011). The overlapping region ($q = 0.12$–1.8 Å^{-1}) served to merge these two data sets, using the in-house Python script-based software developed at X9 (*pyXS*; Allaire & Yang, 2011).

2.7.2. SAXS data analysis. Data analysis was carried out using the in-house *pyXS* software (Allaire & Yang, 2011) and the SAXS program 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 the wider angles, it was decided not to include these data for subsequent analysis, and only the data from the low-angle scattering ($q = 0.017$–0.185 Å^{-1}) were used for the final scattering curve (Fig. 2*a*). The Guinier plot for Axe2-WT ($qR_g = 0.65$–1.26) showed good linearity (adjusted $R^2 = 0.9996$; Fig. 2*b*), 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 38.2 ± 0.1 Å and I_0 to be 3589 ± 5. The *PRIMUS* software (Konarev *et al.*, 2003) was used for data truncation and generation of the pair-distance distribution function $P(r)$ shown in Fig. 2(*c*).

2.7.3. Modelling of the SAXS data. Ten independent *ab initio* models for the molecular envelope of Axe2 were constructed from the $P(r)$ function using *DAMMIN* (Svergun, 1999). The models were generated with no imposition of structure or symmetry, and show an excellent fit to the experimental data (Fig. 2*a*), with χ^2 values ranging between 0.9103 and 0.9192 for each of the ten models. As can be seen in Supplementary Fig. S3, the resulting *ab initio* models provide an unequivocal confirmation of the octameric doughnut-shaped structure of Axe2 obtained from the crystallographic analysis, all showing (to various degrees of definition) the overall circular assembly and central cavity of Axe2.

Improvements to these initial models were made using *DAMAVER* (Volkov & Svergun, 2003), in which the ten *ab initio* models were averaged and filtered to yield the final molecular-envelope model for Axe2-WT presented below. Superimposition of the SAXS-based models with the crystallographic atomic structures was performed with *SUPCOMB* (Kozin & Svergun, 2001), and fitting of the experimental SAXS curves to theoretical curves calculated from crystal structures was performed using *CRYSOL* (Svergun *et al.*, 1995).

2.8. Calculations and figure preparation

The matrices for the superposition of the different structures were calculated by a least-squares distance-minimization algorithm (*LSQ*, implemented within *Coot*). All the figures of the octamer were prepared using PDB coordinates obtained from the *PISA* web server (Krissinel & Henrick, 2007). Figs. 1, 3(*a*), 4(*a*), 5, 6, 8(*a*), 9, 10 and 11 were prepared using *UCSF Chimera* (Pettersen *et al.*, 2004), Figs. 7(*b*) and 7(*c*) using *PyMOL* (v.1.5.0.4; Schrödinger), Figs. 3(*b*) and 4(*b*) using the *Microsoft Office* program *PowerPoint* 2007 and Fig. 2 using *Origin* (OriginLab, Northampton, Massachusetts, USA).

2.9. PDB accession codes

The atomic coordinates of Axe2-WT, Axe2-Se and Axe2-S15A have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (Berman *et al.*, 2000) under accession codes 3w7v, 4jhl and 4jko, respectively.

3. Results and discussion

3.1. The Axe2 monomer

The structure of Axe2-WT is the most relevant for functional conclusions and it will therefore be used as the main basis for the detailed structural analysis presented below. The overall structure of the Axe2 protein corresponds to the

SGNH hydrolase fold. In principle, this fold is similar to the common α/β-hydrolase fold; however, the location of the catalytic residues and the absence of a 'nucleophilic elbow' differentiates the two folds from each other (Wei *et al.*, 1995), as is also the case in the current structure. The fold usually consists of five central parallel β-sheets flanked by two layers of helices. In the case of Axe2, the two helix layers comprise eight α-helices and five 3_{10}-helices (Fig. 3 and Supplementary Fig. S4). The β-sheets of Axe2 are formed by residues 8–13 ($\beta1$), 55–58 ($\beta2$), 84–88 ($\beta3$), 130–134 ($\beta4$) and 168–170 ($\beta5$)

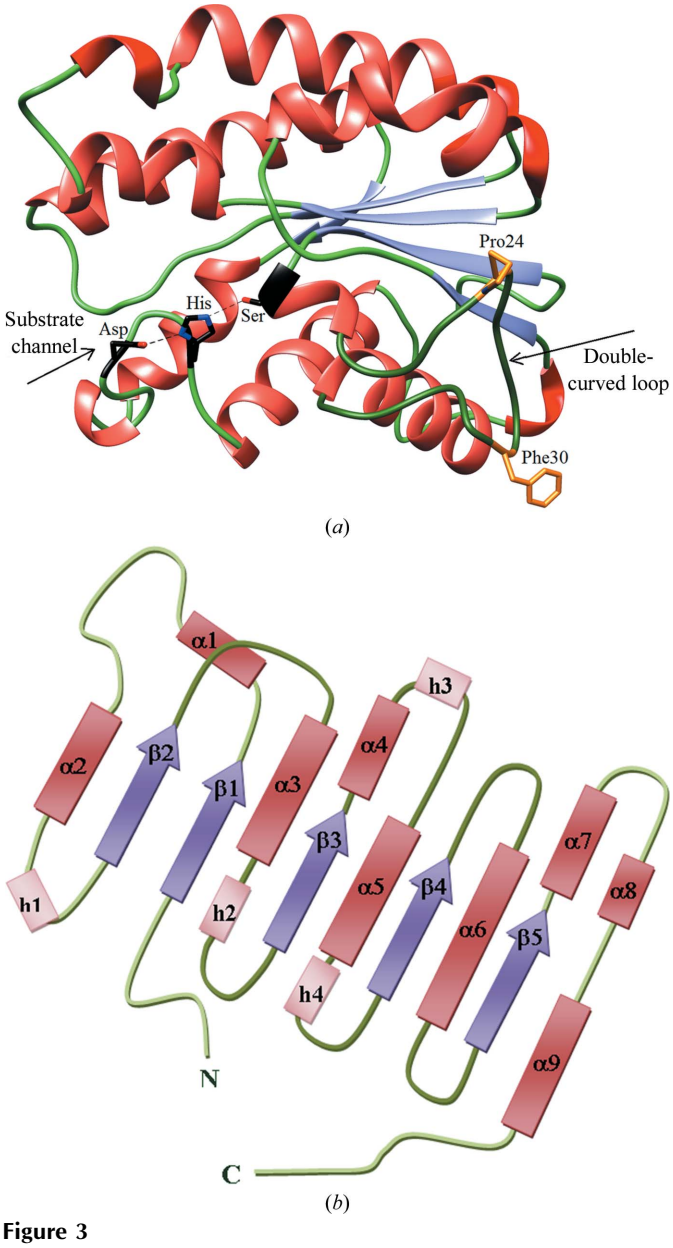


Figure 3
The overall structure of the Axe2 monomer, showing the canonical SGNH hydrolase fold. The central five parallel β-sheets are sandwiched by two layers of α-helices. The α-helices are shown in red, the β-sheets in purple and the loops in green. (*a*) View from the side, where the division into the convex and concave surface is observed. The catalytic triad (marked in black) is seen situated along a cleft in the concave surface and the substrate channel is indicated by an arrow. Pro24 (next to the *cis*-peptide) and Phe30, which are located on the two curves of the 'double-curved loop', are shown in orange. (*b*) A schematic topology diagram of Axe2. The secondary-structure elements are presented and numbered: α-helices (αn) and 3_{10}-helices (hn) as rectangles, β-strands (βn) as arrows and loops as curved single lines.

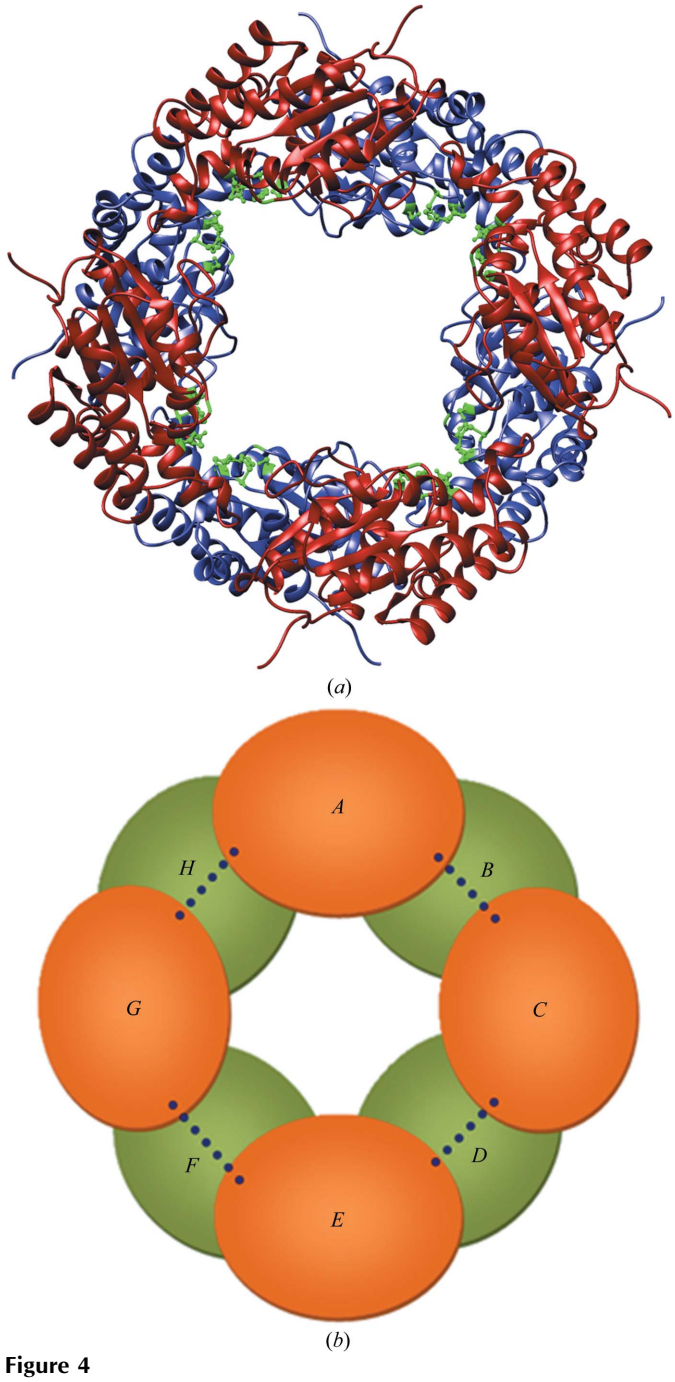


Figure 4
The molecular structure of the Axe2 octamer. (*a*) Ribbon representation of the octameric 'doughnut-shaped' configuration of Axe2, viewed from above. The two tetramers composing the octamer (shown in red and purple) are situated one on top of the other in a staggered manner. The inward-facing catalytic sites (in green) are arranged around the central cavity. (*b*) A scheme of the staggered tetramer layers of Axe2 as viewed from the top of the central cavity. Orange ovals correspond to monomer subunits of one tetramer layer (chains *A*, *C*, *E* and *G*) and green ovals correspond to monomer subunits of the second tetramer layer (chains *B*, *D*, *F* and *H*). Blue dotted lines correspond to the salt-bridge interactions (formed by Arg55 and Glu105) connecting the monomers together within the tetramer.

(Fig. 3*b*), they display a topology of −1x +2x +1x +1x and divide the monomer into two parts: a 'convex surface' and a 'concave surface' (Fig. 3*a*). The α-helices surrounding the β-sheets are formed by residues 14–19 (α1), 36–49 (α2), 65–77 (α3), 90–100 (α4), 109–125 (α5), 145–166 (α6), 171–181 (α7) and 196–211 (α8). The corresponding 3_{10}-helices are formed by residues 50–53 (h1), 78–81 (h2), 103–107 (h3), 126–128 (h4) and 184–188 (h5). Of these helices, α3, α4, α5, α6, h2, h3 and h4 are located on the convex surface of the β-sheets, while α1, α2, α7, α8, h1 and h5 are located on their concave surface.

The catalytic residues of Axe2 were originally determined based on sequence homology (Alalouf *et al.*, 2011). These residues, Ser15, His194 and Asp191, are lined up along a cleft situated on the concave surface of the monomer (Fig. 3*a*). This cleft forms, together with the adjacent amino acids, the substrate channel (Lee *et al.*, 2012). Ser15 is situated on α1, while Asp191 and His194 are situated on a loop between α8 and α9. Thus, Ser15 takes the innermost position, His194 the middle position and Asp191 the position nearest to the protein surface (Fig. 3*a*). A relatively large 'double-curved' loop, formed by residues 19–36, is located near the active site of Axe2. One curve of the loop is formed by the *cis*-peptide bond between residues Arg23 and Pro24, while the other curve is formed by residues Gly28, Ser29 and Phe30 (Fig. 3*a* and Supplementary Fig. S4). The conformational flexibility and functional significance of this loop are discussed further below.

3.2. The Axe2 octamer

As discussed above, the asymmetric unit of the current crystal structure of Axe2 is a dimer, consisting of two Axe2 chains labelled *A* and *B*. Nevertheless, gel-filtration results demonstrate that the Axe2 protein is an octamer in solution (Supplementary Fig. S2), which is probably the biologically relevant form of the enzyme (see below). A closer look at the content of the crystallographic unit cell shows that such an octamer is also present in the current crystal owing to the tetragonal crystal system (space group *I*422), in which the octamer is formed by four symmetry-related asymmetric unit dimers situated around the fourfold axis of the unit cell (Supplementary Fig. S5). To distinguish between the different monomers forming the octamer, the additional Axe2 chains that are symmetry-related to the asymmetric unit dimer (chains *A* and *B*) were labelled *C* and *D*, *E* and *F*, and *G* and *H*, respectively. This highly symmetrical octamer appears as a 'doughnut-shaped' torus built of two circular staggered layers of tetramers situated one over the other (Fig. 4*a*). One tetramer is composed of the chains labelled *A*, *C*, *E* and *G*,

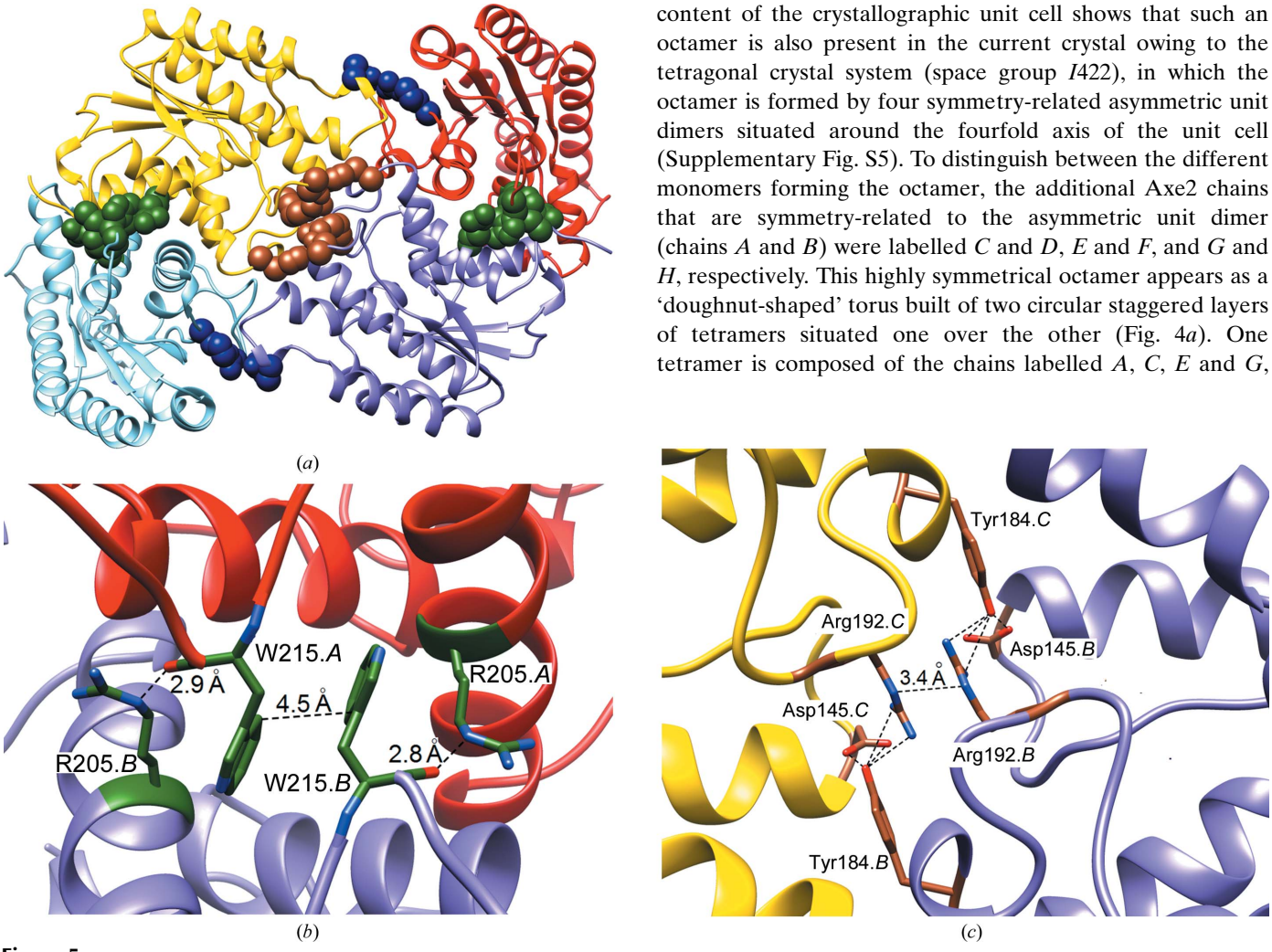


Figure 5
The contacts holding the Axe2 octamer together. (*a*) Ribbon representation of a side view of Axe2 in the same general view as in Supplementary Fig. S6. Chains *A*, *B*, *C* and *D* are shown in red, purple, yellow and light blue, respectively. The residues forming the main connections between these chains are shown in a space-filling representation. Specifically, the residues forming the asymmetric unit dimers are in green, the salt bridge forming the tetramers is in dark blue and the hydrogen-bond cluster connecting the four asymmetric unit dimers is in brown. (*b*) Close-up of the contacts between the monomers of the crystallographic asymmetric unit forming the asymmetric unit dimers, including the π-stacking between two Trp215 residues and the hydrogen bonds between two pairs of Trp215 and Arg205 residues. (*c*) Close-up into the central intermolecular hydrogen-bond cluster involving Asp145, Tyr184 and Arg192, where the key role of Tyr184 is demonstrated.

while the other is composed of the chains labelled *B*, *D*, *F* and *H* (Fig. 4*b* and Supplementary Fig. S6). The chains of the tetramer are related to one another like chains *A* and *B* of the crystallographic asymmetric unit. Superposition of the two tetramers shows an almost complete overlap between them (r.m.s.d. of 0.592 Å based on 876 C^{α} atoms, r.m.s.d. of 0.608 Å based on all main-chain atoms and r.m.s.d. of 0.891 Å based on all atoms).

The monomer–monomer contacts forming the 'asymmetric unit dimers' (*i.e.* chains *A* and *B*, chains *C* and *D*, chains *E* and *F*, and chains *G* and *H*) are mainly based on three significant interactions, two nonconserved hydrogen bonds and a π-stacking interaction, all between residues located in the C-terminal region of the two protein chains and in the centre of the monomer–monomer interface (Fig. 5*a*, green contacts). Specifically, looking at the 'asymmetric unit dimer *AB*', the π-stacking interaction is between the aromatic side chains of two Trp215 residues, one from chain *A* and one from chain *B*. The two hydrogen bonds are formed in the same general area, between Arg205.*A* NH2 of molecule *A* and Trp215.*B* O of molecule *B* and similarly (and almost symmetrically) between the reverse pair Arg205.*B* NH2 and Trp215.*A* O (Fig. 5*b*). The relative strength and significance of these interactions are further discussed below.

Each tetramer composing the octamer (chains *A*, *C*, *E* and *G* in tetramer I, and chains *B*, *D*, *F* and *H* in tetramer II) is formed by four conserved salt-bridge contacts between Glu105 OE1 and OE2 of one chain and Arg55 NH1 and NH2 of the nearby chain [Figs. 4*b* and 5*a* (blue contacts) and Supplementary Fig. S7]. The interactions forming the tetramer–tetramer interface to generate the full octameric structure include the contacts forming the 'asymmetric unit dimers' (as stated above; Fig. 5*a*, green contacts) and a cluster of hydrogen bonds and π-stacking interactions connecting these four 'asymmetric unit dimers' together (*i.e.* links between chains *B* and *C*, chains *D* and *E*, chains *F* and *G*, and chains *H* and *A*), contacts situated in the centre of the protein near the catalytic sites (Fig. 5*a*, brown contacts).

The cluster of hydrogen bonds consists of contacts between Tyr184 OH of one chain and Arg192 NH1 and NH2 and Asp145 OD1 and OD2 of another. Specifically, looking at the 'dimers' *AB* and *CD*, the connections between these two dimers are formed between Tyr184.*B* OH of molecule *B* and Arg192.*C* NH1 and NH2 and Asp145.*C* OD1 and OD2 of molecule *C*, and similarly between Tyr184.*C* OH and Arg192.*B* NH1 and NH2 and Asp145.*B* OD1 and OD2. It is noted that the secondary nitrogen atoms of Arg192.*B* NH3 and Arg192.*C* NH3 are separated from one another by about only 3.4 Å, and thus the two sets of hydrogen bonds are in close proximity to one another, forming a very strong cluster of hydrogen bonds (Fig. 5*c*). The π-stacking interactions take place between the side chains of Trp95.*B* of molecule *B* and Trp190.*C* of molecule *C*, and similarly between Trp190.*B* and Trp95.*C*. The main intermolecular interactions involved in formation and stabilization of the dimers, the tetramers and the octamer of the current structure of Axe2 are summarized in Table 3.

Table 3
Summary of intermolecular contacts involved in the Axe2 octamer.

	Type of contact	Residues in contact	Atoms in contact	Distance (Å)
Contacts forming 'tetramers'	Salt bridge	Glu105.*B*–Arg55.*D* Glu105.*D*–Arg55.*F* Glu105.*F*–Arg55.*H* Glu105.*H*–Arg55.*B*	OE1–NH1 OE2–NH2	2.59 2.84
	Salt bridge	Glu105.*A*–Arg55.*G* Glu105.*G*–Arg55.*E* Glu105.*E*–Arg55.*C* Glu105.*C*–Arg55.*A*	OE1–NH1 OE2–NH2	2.92 3.37
Contacts within 'asymmetric unit dimers'	Hydrogen bond	Arg205.*A*–Trp215.*B* Arg205.*C*–Trp215.*D* Arg205.*E*–Trp215.*F* Arg205.*G*–Trp215.*H*	NE–O	2.88
	Hydrogen bond	Trp215.*A*–Arg205.*B* Trp215.*C*–Arg205.*D* Trp215.*E*–Arg205.*F* Trp215.*G*–Arg205.*H*	O–NE	2.88
	π–Stacking	Trp215.*A*–Trp215.*B* Trp215.*C*–Trp215.*D* Trp215.*E*–Trp215.*F* Trp215.*G*–Trp215.*H*	CB–CB	3.54
Contacts forming the octamer	Hydrogen bond	Tyr184.*B*–Asp145.*C* Tyr184.*D*–Asp145.*E* Tyr184.*F*–Asp145.*G* Tyr184.*H*–Asp145.*A*	OH–OD1 OH–OD2	3.17 2.72
	Hydrogen bond	Tyr184.*B*–Arg192.*C* Tyr184.*D*–Arg192.*E* Tyr184.*F*–Arg192.*G* Tyr184.*H*–Arg192.*A*	OH–NH1 OH–NH2	3.39 3.24
	Hydrogen bond	Asp145.*B*–Tyr184.*C* Asp145.*D*–Tyr184.*E* Asp145.*F*–Tyr184.*G* Asp145.*H*–Tyr184.*A*	OD1–OH OD2–OH	3.18 2.56
	Hydrogen bond	Arg192.*B*–Tyr184.*C* Arg192.*D*–Tyr184.*E* Arg192.*F*–Tyr184.*G* Arg192.*H*–Tyr184.*A*	NH1–OH NH2–OH	3.33 3.34
	π-Stacking	Trp95.*B*–Trp190.*C* Trp95.*D*–Trp190.*E* Trp95.*F*–Trp190.*G* Trp95.*H*–Trp190.*A*	CH2–CH2	3.59
	π-Stacking	Trp190.*B*–Trp95.*C* Trp190.*D*–Trp95.*E* Trp190.*F*–Trp95.*G* Trp190.*H*–Trp95.*A*	CH2–CH2	3.65

The eight polypeptide chains of the Axe2 octamer are situated around a central cavity, which can be compartmented into four local 'hollows'. Interestingly, two active sites are situated in each of these hollows (active sites that belong to different chains and different tetramers), separated from one another only by approximately 5 Å (Fig. 6*a*). Thus, the enzyme possesses four pairs of 'double' catalytic sites, situated around the C_4 axis of the octamer, all facing the central cavity formed in its centre. Interestingly, as discussed in §3.3 below, each of these active sites are further stabilized by the nearby hydrogen-bond cluster (Fig. 6*b*) and π-stacking interactions mentioned above. The overall size of the octamer is about 100 Å in diameter, while the cavity is about 35 Å in diameter. This is a relatively large cavity, as discussed later. The general molecular shape and dimensions are further confirmed by TEM experiments, which were performed on similarly prepared samples of the Axe2-WT protein. The resulting

TEM images demonstrate a number of circular proteins, all around 100 Å in diameter (Fig. 7*a*), correlating well with the shape and dimensions of the crystallographic octamer presented above.

Recently collected small-angle X-ray scattering (SAXS) data provide further evidence for the torus-like octameric form of Axe2. Independently of the crystal structure, and with no imposition of symmetry, the SAXS data produced a circular molecular envelope showing a remarkably good fit when superimposed upon the atomic crystallographic structure presented above. It should be noted that the good fit applies not only to the general circular structure of Axe2 but also to its central cavity and some of the secondary-structure elements on its surface (Figs. 7*b* and 7*c*). In addition, comparing the experimental scattering curve of Axe2 {log[$I(q)$] *versus* q} with the simulated scattering curve obtained from its crystallographic model (calculated with *CRYSOL*), a relatively good fit is demonstrated (Fig. 2*a*). Except for some minor deviations, mainly at larger values of q (0.15–0.18), the simulated curve is well within the error range of the experimental data, serving as further evidence for the reliability of the SAXS data and the resulting molecular envelope.

The very good agreement of the crystallographic model with the SAXS data is especially significant, as the protein solution used for the SAXS data collection is more similar to the physiological environment of the protein than the environment in protein crystals, where potential conformational changes and nonbiological interactions are possible. Hence, it appears as if the specific Axe2 octamer described above is indeed the main form existing in solution and as such is very likely to represent the relevant and functional biological form of the Axe2 protein.

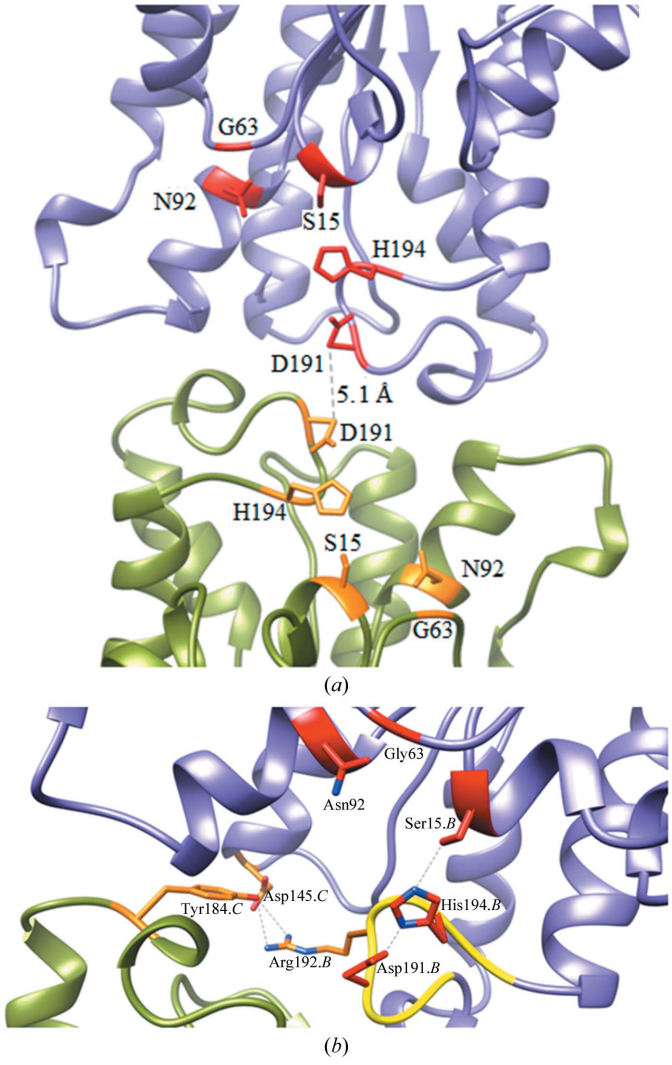


Figure 6
The active site of Axe2. (*a*) Ribbon representation of two monomer subunits from different tetramers of Axe2, showing two active sites in close proximity (5.1 Å) to one another forming a 'pair' of catalytic sites. One monomer is shown in purple with its active site in red (top). The other monomer is shown in green with its active site in orange (bottom). (*b*) The hydrogen-bond cluster formed by Asp145, Tyr184 and Arg192, showing its stabilizing effect on the Asp191–His194 loop (yellow). The loop conformation positions both Asp191 and His194 in the correct catalytic orientation with respect to Ser15.

3.3. Possible significance of the Axe2 octameric structure

The doughnut-shaped octameric structure of Axe2, as described above, is rather unusual. Several cases of hexameric doughnut-shaped hydrolases have been reported in the past (*e.g.* Hövel *et al.*, 2003; Vincent *et al.*, 2003; Montoro-García *et al.*, 2011), but as far as we know this is the first case of an octameric doughnut-shaped assembly of a hydrolase with inward-facing active sites (see below). One possible reason for this unique assembly is further stabilization around the active site. As described above, the cluster of hydrogen bonds and π-stacking interactions in the centre of the enzyme are located near the active sites. These contacts may play a role in assisting Asp191 and His194 to maintain the correct orientation with respect to Ser15. The contacts between Arg192 and Tyr184 (Fig. 6*b*) and between Trp190 and Trp95 help to bring the loop carrying Asp191 and His194 into proximity to Ser15 with the correct position and orientation, forming the proper geometry of the Asp-His-Ser catalytic triad.

Another possible reason for this unique structure is to sequester and shield the active sites away from the general cytoplasmic contents and thus 'protect' the substrate from unintentional and unregulated hydrolysis (Vincent *et al.*, 2003). As stated above, all the active sites of the current Axe2 structure face the central cavity (Fig. 4*a*) and as such are exposed only to the selected contents of the internal cavity formed in the centre of this torus octamer. A similar argument may be drawn in which the relatively limited central cavity may function to hinder large substrates from entering the central space and reaching the surrounding active sites. Such size selection will enhance the specificity of the otherwise rather promiscuous active centre by allowing only substrates of the correct size and shape to enter the central cavity and the active sites (Vincent *et al.*, 2003; Montoro-García *et al.*, 2011).

Improvement of the catalytic activity may be another explanation for the unique multimeric structure of the Axe2 octamer. Owing to this assembly, two active sites are brought

into close proximity to one another (Fig. 6*a*), and this may provide various synergistic effects to catalysis. For example, in cases where the hemicellulose substrate is decorated with a number of acetyl groups, this special assembly may allow the two active sites to work simultaneously on the same substrate, thus doubling the rate of overall hydrolysis. In this regard, the enzyme has been shown to remove two acetyl groups simultaneously from positions 3 and 4 of 2,3,4-tri-*O*-acetyl-methyl-β-D-xylopyranoside (Alalouf *et al.*, 2011).

The most trivial, and probably the most likely, explanation for the special octameric assembly observed in the current structure, however, is its contribution to the overall stability of the enzyme. As an octamer, there are significantly more intermolecular interactions compared with a dimer, a tetramer or a hexamer of the same protein, leading to a corresponding enhancement of the thermodynamic stability of its multimeric assembly.

Obviously, although each of these different hypotheses could be relevant, their validity should be tested and confirmed by further specific, carefully designed, experiments. Such studies are currently under way in our laboratory. In this regard, it should be noted that despite the large volume of experimental and theoretical work, the role of the quaternary structure of enzymes is still far from being understood (Devenish & Gerrard, 2009; Griffin & Gerrard, 2012; Matthews & Sunde, 2012), as will be further discussed below.

3.4. The active site

The active site of Axe2 consists of the catalytic triad typical of serine proteases and serine esterases. Here, the specific catalytic residues are Ser15, His194 and Asp191 (Alalouf *et al.*, 2011), which are positioned in the proper orientation for the canonical charge-relay system mechanism (Banacký & Linder, 1981; Figs. 3*a* and 6*b*). The three functional groups are connected by two relatively strong hydrogen bonds: one from Ser15 OG to His194 NE2 (2.9 Å) and the other from His194 ND1 to Asp191 OD2 (2.7 Å). These interactions confirm the role of Asp191 and His194 as amplifiers of the nucleophilicity of Ser15, promoting the nucleophilic attack of the deprotonated Ser15 side chain on the target ester bond of the substrate as the first step in its hydrolysis. These structural conclusions are fully supported by the corresponding biochemical experiments, which demonstrate that all three catalytic mutants (S15A, H194A and D191A) are practically inactive (Alalouf *et al.*, 2011).

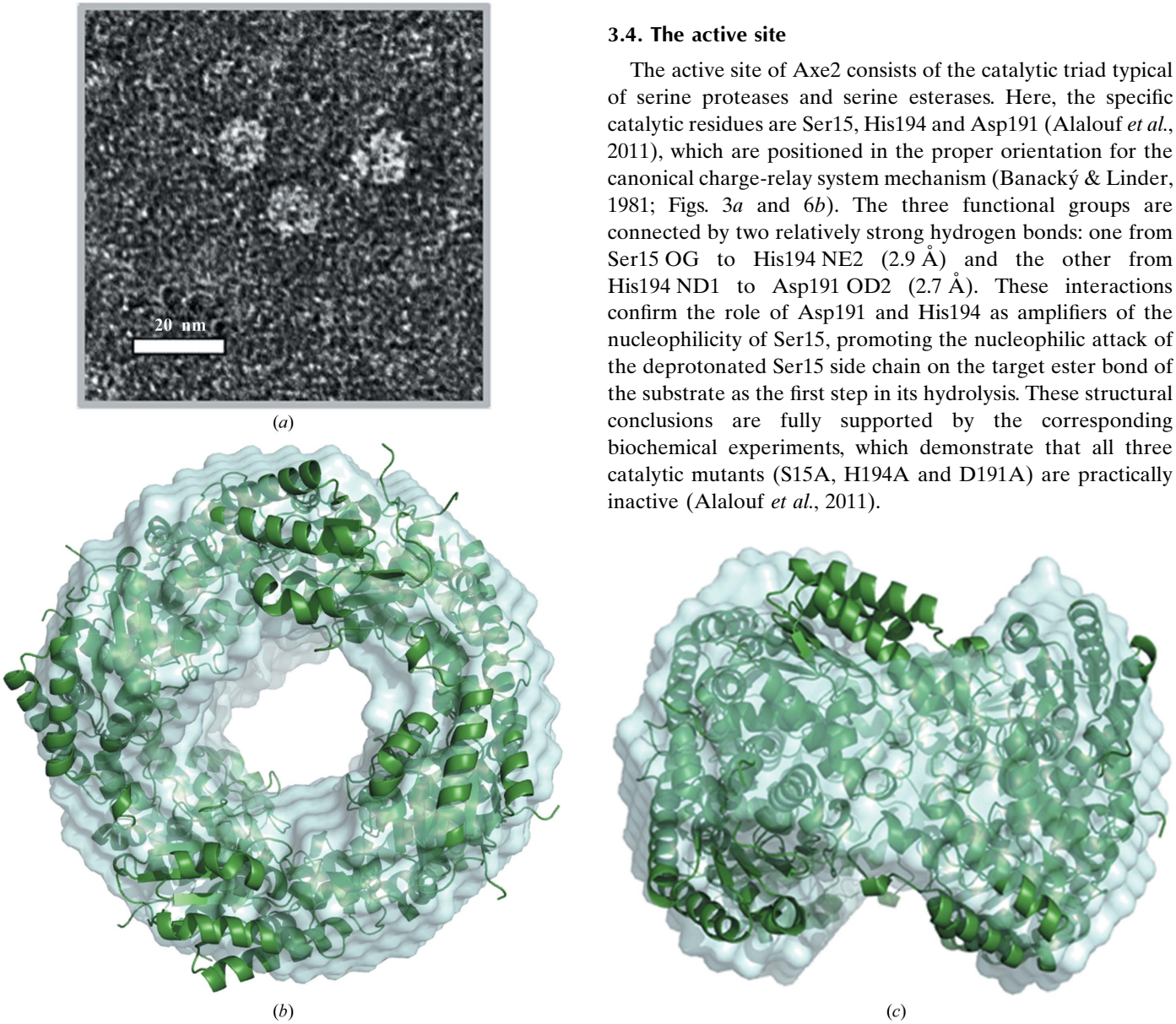


Figure 7
Confirmation of the crystallographic quaternary structure of Axe2 by alternative methods. (*a*) A TEM image of Axe2 in solution, demonstrating three circular proteins approximately 10 nm in diameter (centre). (*b*) Superposition of the molecular envelope obtained from SAXS data on the crystallographic model of Axe2, demonstrating a very good fit between the SAXS and the crystallographic results and confirming the octameric circular assembly of Axe2 in solution. (*c*) A view from the side of the superposition of the SAXS molecular envelope upon the crystallographic model.

Based on homology to other members of the SGNH hydrolase family, the residues forming the oxyanion hole were suggested to be Gly63 and Asn92 (Alalouf *et al.*, 2011). Our current findings confirm these roles, as the main-chain amine groups of both residues are at the correct distance and in the correct orientation to interact and stabilize the anionic catalytic intermediate formed in the active site (Fig. 6*b*). Further support for this comes from the current structure of the Axe2-S15A catalytic mutant, in which a phosphate anion was found bound in the active site of the enzyme (see above). In this case, specific interactions with the phosphate oxygen atoms are formed by His194 NE2, Gly63 N, Asn92 OD1 and ND2, Ser15Ala N and HOH647 (Fig. 8), and the bound phosphate anion seems to serve as a perfect mimic for the true tetrahedral intermediate of the catalytic reaction. Moreover, in the other structural models of Axe2 presented here, chloride anions were found in the same site instead. These anions are similarly coordinated to His194 NE2, Gly63 N, Asn92 OD1 and ND2, and Ser15 OG and N. The fact that both PO_4^{3-} and Cl^- were found to be strongly bound in the active site and to specifically interact with Gly63 and Asn92 confirms the involvement of these amino acids in the formation of the oxyanion hole, as the suggested role of this hole is to stabilize the negative charge formed by the acyl-enzyme tetrahedral intermediate.

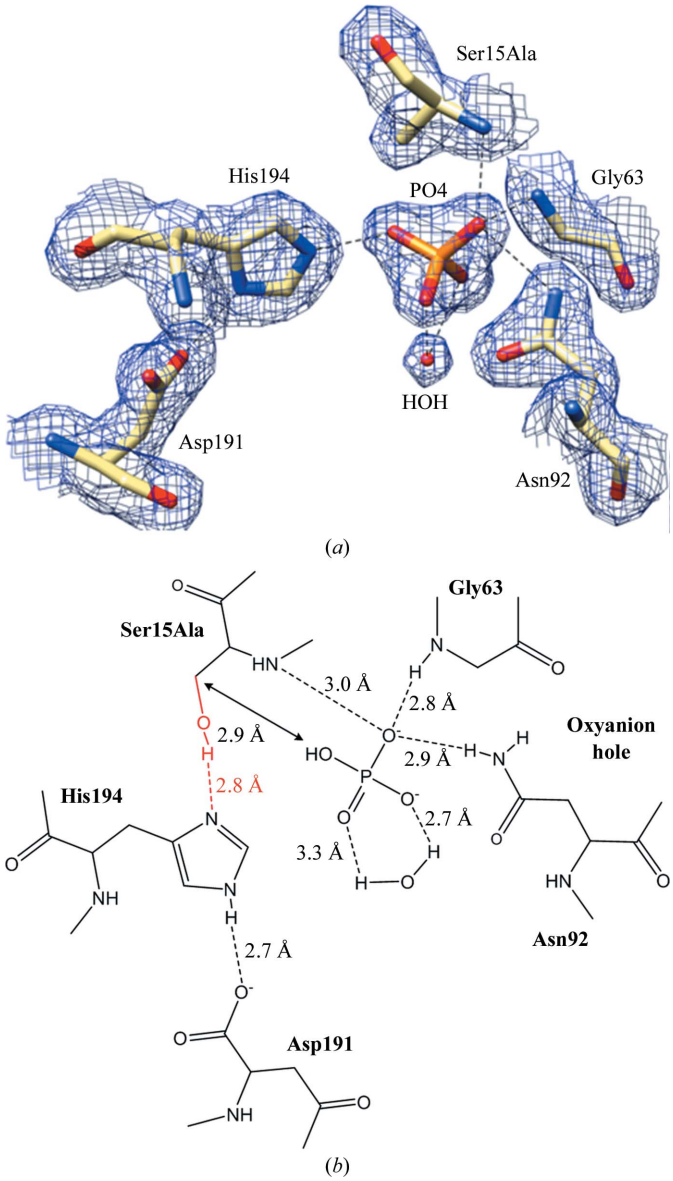


Figure 8
The active site of the nucleophile catalytic mutant Axe2-S15A. (*a*) A section of the electron-density map of Axe2-S15A superimposed on the final model of the active site. The hydrogen bonds connecting the bound phosphate ion (PO4) to the oxyanion hole and the catalytic residues are marked by dotted lines. The map is at 1.90 Å resolution and is contoured at the 1.5σ level. (*b*) A scheme of the active site of Axe2-S15A, showing the interactions formed by the bound phosphate ion, including the relevant distances. Superimposed on this scheme is a modelled hydroxyl group of the WT Ser15, including its calculated distances from the phosphate ion and His194 (coloured red).

Although the active site of each of the Axe2 monomers is quite standard for serine esterases, the oligomeric arrangement of the active sites within the Axe2 octamer is not. This arrangement is unusual since every two active sites in the octameric assembly are in close proximity to one another, thus forming four 'pairs' of active sites (see §3.2; Figs. 4*a* and 6*a*). In comparison, the six active sites in the hexameric esterases reported previously are completely separated from one another (Vincent *et al.*, 2003; Montoro-García *et al.*, 2011), an arrangement that seems to be more common in multimeric enzymes. At this point, we could not find a reasonable functional explanation for these unusual pairs of catalytic triads, except for the possible advantage in the case of substrates with a number of close acetyl substitutions, as discussed above. Moreover, these 'double' active sites are situated so that the substrate channels of every active site do not face each other, but instead are rotated to the opposite direction and translated by about 6–9 Å with respect to one another (Fig. 6*a*). This is counterintuitive because two substrate channels facing each other would suggest simultaneous activity of the active-site pair on the same substrate, potentially achieving a synergistic effect, while this 'back-to-back' arrangement does not appear to enable such an effect. The functional explanations for this seemingly less efficient oligomeric arrangement, if any, will hopefully become clearer upon further experiments.

3.5. Special structural features

Several glycerol molecules (originating from the glycerol-rich cryoprotecting solvent) were introduced into the electron-density map and are part of the final crystallographic models of Axe2. There are five such glycerols in the Axe2-WT structure, seven in Axe2-Se and seven in Axe2-S15A. Most of these solvent molecules appear on the surface of the protein and are bound by at least one hydrogen bond to surface functional groups (charged or polar). All of these bound glycerol molecules do not seem to affect the overall conformation of the protein, especially since they are relatively small and occupy only those areas which are likely to host water molecules in pure aqueous solutions. We thus expect the

structures of the Axe2 models to be practically identical in a glycerol-free (physiological) environment.

As stated above, the Axe2 polypeptide chain contains a single *cis*-peptide bond, which is located between residues Arg23 and Pro24. The position of this *cis*-peptide bond seems to be critical, as it influences the exact orientation of the nearby conserved residue Glu27 (in comparison with other homologous proteins) and allows the formation of a crucial hydrogen bond between its OE2 group and the NE group of Arg55, as demonstrated in Fig. 9(*a*). This hydrogen bond seems to stabilize Arg55 and thus orients it to form the conserved salt-bridge contact with Glu105, which is one of the key interactions holding the Axe2 tetramers together (see §3.2 and Supplementary Fig. S7).

As described above (§2.4), both the glycerol molecules (in the case of Axe2-S15A) and the chloride ions (in the case of Axe2-Se) were found to be bound near the loop formed by Gly28, Ser29 and Phe30. These residues form one curve of the relatively long 'double-curved' loop situated near the active site (see above). The other curve of this loose loop is formed by the *cis*-peptide bond between Arg23 and Pro24 (Fig. 9*b*). Superposition of the three Axe2 structures presented here demonstrates that this loop is quite flexible and moves significantly from one structure to the other, as could also be inferred from the high temperature factors and the widely spread and disordered electron density observed in this area of the protein in the experimental maps. The function of this loop and the reason for its exceptional flexibility, if any, is not fully clear as yet. Nevertheless, the conformational changes of this loop, its location near the active site and the glycerol molecules (Fig. 9*b*) and chloride ions found nearby (in Axe2-S15A and Axe2-Se) suggest that it may be involved in the catalytic function of Axe2, probably in guiding the substrate to its destination. In this respect, the position and conformational flexibility of the aromatic side chain of Phe30 in the present structure of Axe2 seem to be analogous to that of Phe4 in the hexameric structure of the thermostable acetylesterase from *T. maritima* (TM0077; Levisson *et al.*, 2012; see below), which was shown to play a role in the closing and opening of the central cavity of the oligomeric assembly upon substrate binding.

3.6. Comparison with related protein structures

3.6.1. Comparison with a highly homologous protein from the GDSL family. As a member of the GDSL family, the structure of Axe2 may be compared with structures of other available structures of proteins from the same family. Using the *DALI* server (Holm & Sander, 1995), a highly homologous protein (3RJT), which has 52% sequence identity to Axe2, has been found. 3RJT (PDB entry 3rjt; Midwest Center for Structural Genomics, unpublished work) is a putative lipolytic protein from *Alicyclobacillus acidocaldarius* subsp. *acidocaldarius* DSM446 belonging to the GDSL family and containing 213 amino acids per monomer. As expected from such high sequence identity, a superposition of the structures of Axe2 and 3RJT shows a very good overlap (Fig. 10*a*), with an r.m.s.d. value of 0.66 Å for 212 of the 213 residues aligned. In retrospect, this structure could have served as an excellent model for molecular replacement, had it been available at the time of the structure determination of Axe2.

As can be seen from the structural overlap, and as expected for such highly homologous proteins from the same family, the secondary-structure elements of the two proteins are very similar. Nevertheless, a slight difference exists in the number of β-strands in each protein. Axe2 contains five parallel strands, whereas 3RJT contains an additional unique very short antiparallel strand. Some deviations from the more common five central parallel β-strands have been seen before in the SGNH hydrolase superfamily, an example being the case of the enzyme At4g34215 from *Arabidopsis thaliana* (PDB entry 2apj; Bitto *et al.*, 2005), in which six parallel β-strands and one antiparallel β-strand are apparent. This raises questions regarding the exact definition of the relatively newly discovered SGNH hydrolase fold (Mølgaard *et al.*, 2000) and whether all proteins belonging to the SGNH hydrolase superfamily possess this exact fold.

The high structural similarity between Axe2 and 3RJT is reflected in their loops, which correspond rather well in number, location and orientation. This is rather remarkable, as loops are usually more flexible and varied between proteins.

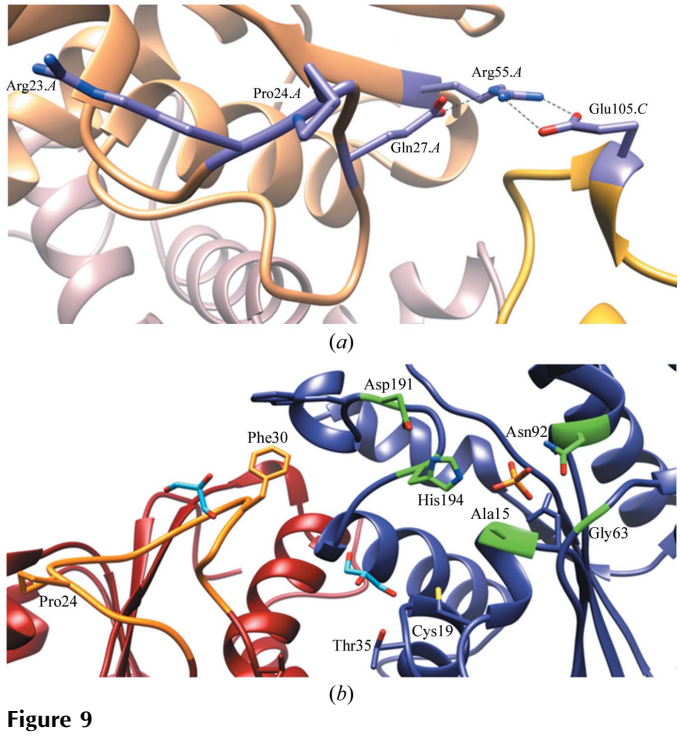
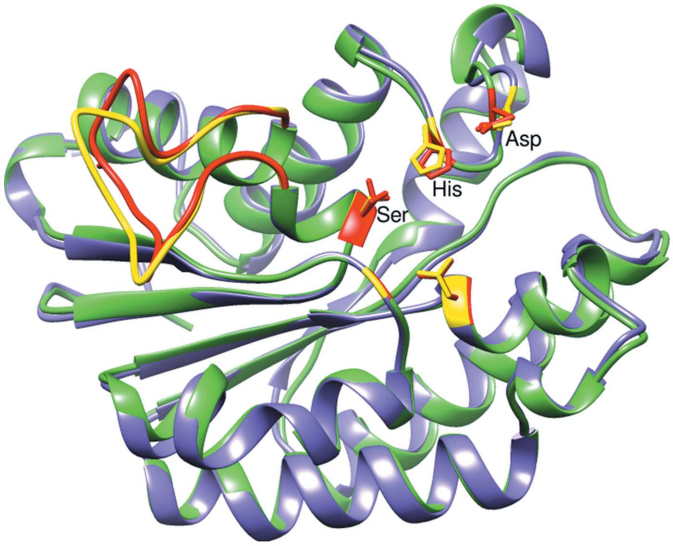


Figure 9
The unique 'double-curved' loop of Axe2. (*a*) The first curve of the 'double-curved' loop, which is facilitated by the *cis*-peptide bond between Arg23 and Pro24 (centre left). The conformation imposed by this *cis*-peptide bond positions Glu27 to interact with and stabilize Arg55, enabling Arg55 to form the conserved salt bridge with Glu105. (*b*) The 'double-curved' loop of Axe2 (orange), as found in the model of Axe2-S15A. This loop may play a role in bringing substrate into the active site (coloured green) of the adjacent Axe2-S15A monomer, as suggested by the glycerol molecules (turquoise) found in its vicinity. Chain *A* of the 'asymmetric unit dimer' is shown in red and chain *B* is shown in purple. The bound phosphate ion is shown in orange (centre right).

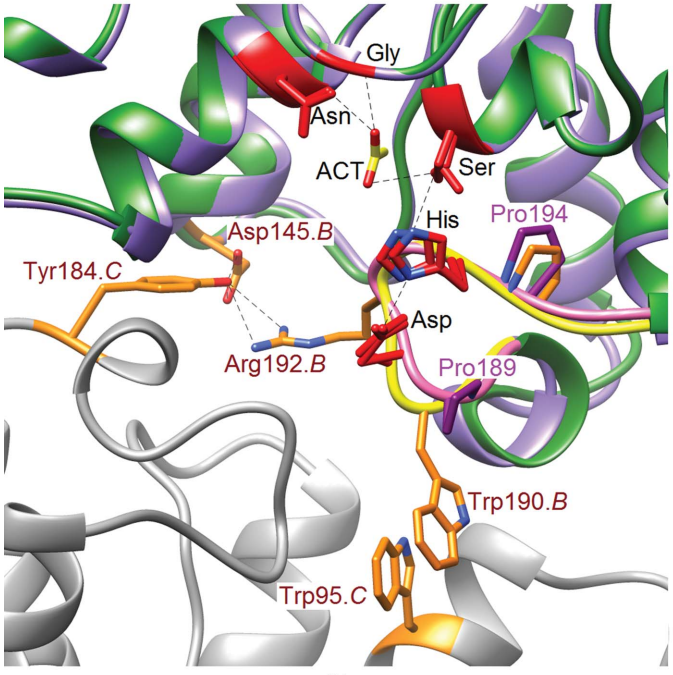
An exception is the large 'double-curved' loop of Axe2 (residues 19–36), which moves noticeably in respect to the analogous loop in 3RJT (residues 18–35; Fig. 10*a*). This conformational movement provides further evidence for the proposed activity and flexibility of this loop in guiding substrate to the catalytic site, as stated above and discussed further below.

As expected, the active sites of Axe2 and 3RJT fit very well (Fig. 10*b*). A superposition of the two structures reveals the previously unidentified catalytic residues of 3RJT, confirming that there is a shift of one amino acid between the catalytic residues of Axe2 with respect to those of 3RJT. Thus, the apparent catalytic triad of 3RJT is made up of Ser14, Asp190 and His193, and the oxyanion hole is made up of Gly62 and Asn91. Interestingly, an acetate molecule (ACT) is observed in a key position in the active site of 3RJT (Fig. 10*b*), which seems to mimic the negatively charged species along the catalytic reaction coordinate of the enzyme. The distance and orientation of this acetate anion with respect to residues Gly62 and Asn91 in the 3RJT structure provide further proof for the role of these amino acids (and the corresponding Gly63 and Asn92 of Axe2) as part of the oxyanion hole.

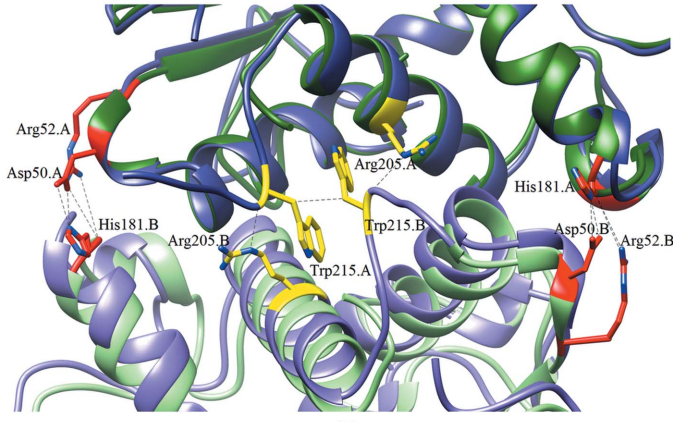
Another interesting difference is observed near the active site, where Trp190 of Axe2 is replaced by Pro189 in 3RJT (Fig. 10*b*). This specific residue difference may be explained, at least in part, in terms of the structural stabilization of the catalytic assembly. As stated above, the hydrogen-bond cluster formed by Asp145, Tyr184 and Arg192, and the π-stacking interactions between Trp190 and Trp95, stabilize the octameric assembly of Axe2. These interactions also stabilize the active site geometry by bringing the loop carrying Asp191 and His194 into the proper orientation with respect to Ser15. Such conformational stabilization should be critical for the correct operation of the catalytic triad of Axe2, thereby enabling the canonical charge-relay mechanism. The analogous residues to Tyr184 of Axe2 are not present in 3RJT and hence this key



(*a*)



(*b*)



(*c*)

Figure 10
Comparison of the Axe2 structure with the structure of 3RJT, a highly homologous protein (52% sequence identity; PDB entry 3rjt). (*a*) Overlap of monomer *A* of Axe2 (purple) with monomer *A* of 3RJT (green). The active site of Axe2 and its 'double-curved' loop (residues 19–36) are shown in yellow and the active site of 3RJT and its analogous loop (residues 18–35) are shown in red. A very good fit is demonstrated between the secondary structures of both proteins, except for minor deviations observed in the coloured 'double-curved' loops. (*b*) Overlap between the active sites of chains *B* of Axe2 (purple) and 3RJT (green). The residues composing the active sites are shown in red for both proteins. The loops carrying the catalytic Asp and His are highlighted for Axe2 (yellow) and 3RJT (pink), showing how stabilization of this key loop is achieved in both cases. In Axe2 it is held in position by the hydrogen-bond cluster involving Asp145.*B*, Arg192.*B* and Tyr184.*C*, and the π-stacking interaction between Trp190.*B* and Trp95.*C* (in orange; chain *C* in grey), whereas in 3RJT, it is conformationally stabilized by Pro189 and Pro194 (in bright purple). In addition, a bound acetate molecule (ACT, in yellow) may be seen in the active site of 3RJT. (*c*) Superposition of the 3RJT dimer (green) with the 'asymmetric unit dimer' of Axe2 (purple), zooming into the monomer–monomer interface of both proteins. It is noted that despite the high structural similarity, the two proteins differ significantly in the specific interactions holding these dimers together. In Axe2 there are three contacts (one aromatic stacking and two hydrogen bonds) located at the centre of the interface (shown in yellow). In 3RJT there are four tight hydrogen bonds located at the two ends of the interface (shown in red).

conformational stabilization is absent. It seems that this handicap is partially overcome by two proline residues in 3RJT, Pro189 and Pro194, which probably rigidify the corresponding loop carrying the catalytic Asp and His, thus stabilizing this critical loop in the correct conformation in an alternative manner (Fig. 10*b*). Obviously, this interesting structure–function argument should be further examined by specific site-directed experiments, as suggested below.

A somewhat puzzling question concerns the specific interactions that stabilize the crystallographic dimeric structure in the case of 3RJT compared with the corresponding interactions in Axe2. A series of least-squares structural comparisons between the crystallographic dimer reported for 3RJT (PDB entry 3rjt) and the three types of dimers observed in the current crystallographic octamer of Axe2 (*A*–*B*, *A*–*C* and *B*–*C*) demonstrate clearly that the 3RJT dimer is highly structurally homologous to the 'asymmetric unit dimer' (*A*–*B*) of Axe2 (Fig. 10*c*), with an r.m.s.d. value of 0.99 Å for 421 of the 428 residues aligned. Surprisingly, this high overall structural similarity does not extend to the key intermolecular contacts holding these two dimers together. As described in §3.2 above, in the Axe2 structure the main contacts forming the 'asymmetric unit dimers' include the π-stacking interaction between two Trp215 residues (from chains *A* and *B*), as well as two hydrogen bonds: one between Arg205.*A* NH2 and Trp215.*B* O and the second between the reverse pair Arg205.*B* NH2 and Trp215.*A* O (Fig. 5*b*). These three interactions are located at the centre of the *A*–*B* monomer–monomer interface, in which the end of the C-terminus of the two interacting polypeptides is located (Fig. 10*c*). Interestingly, in 3RJT this polypeptide part is completely missing (3RJT contains 213 amino acids *versus* the 219 amino acids of Axe2) and none of these three interactions take place. Alternatively, in 3RJT the key interactions are located at the two sides of the monomer–monomer interface and include four tight hydrogen bonds, with the first two being between His181.*A* ND1 and Asp50.*B* OD1 and OD2, and between His181.*A* O and Arg52.*B* NH1 and NH2 at one end of the interface. The other two hydrogen bonds are between their symmetrical counterparts (His181.*B* ND1 and Asp50.*A* OD1 and OD2, and His181.*B* O and Arg52.*A* NH1 and NH2) at the other end of the interface (Fig. 10*c*). This striking difference is rather unexpected considering the high sequence and structural homology between these otherwise very similar proteins. One of the possible explanations for this could be related, at least partly, to the fact that in Axe2 the 'asymmetric unit dimer' is part of a larger octameric assembly and as such is involved in further interactions with neighbouring monomers (*A*–*C* and *B*–*C*). These extra interactions, which are apparently absent in the reported crystallographic dimer of 3RJT, could change both the type and the position of the contacts needed to stabilize the 'asymmetric unit dimer', leading to the alternative monomer–monomer interactions observed in Axe2.

A final important point of comparison concerns the oligomeric structure of 3RJT. No experimental data have yet been published regarding this protein; however, according to its PDB file (PDB entry 3rjt) 3RJT is a dimer in solution based on calculations by the *PISA* server (Krissinel & Henrick, 2007). This statement is in contrast to

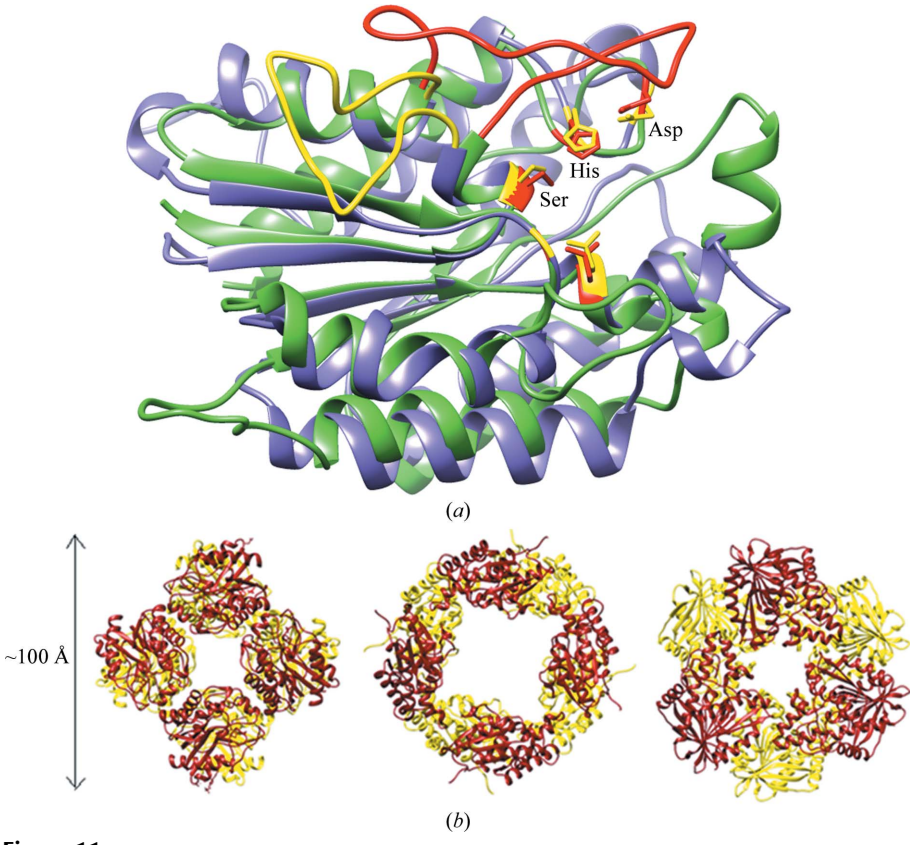


Figure 11
Comparison of the Axe2 structure with the structure of the related octameric hydrolase MsAcT (PDB entry 2q0q). (*a*) Overlap of monomer *A* of Axe2 with monomer *A* of MsAcT, showing the overall similarities between the secondary structures of the proteins. The Axe2 monomer is shown in purple and MsAcT is shown in green. The active site of Axe2 and its 'double-curved' loop (residues 19–36) are shown in yellow and the active site of MsAcT and its analogous loop (residues 15–33) are shown in red. A significant difference in the conformation of these loops is demonstrated. (*b*) Comparison of the crystallographic octameric structure of Axe2 to those of two other ring-like hydrolases: left, MsAcT (octamer); middle, Axe2 (octamer); right, BpAXE (hexamer). The overall size of these multimers is around 100 Å (a scale bar is shown on the left). The two layers composing these ring-like structures are shown with the top layer coloured red and the bottom layer coloured yellow. It can be seen that the layers in Axe2 and BpAXE are arranged in a staggered manner, while the layers in MsAcT are arranged in an eclipsed manner. It is also demonstrated that the relative size of the central cavity of Axe2 is significantly larger compared with those in the other two structures.

the octameric form of Axe2 presented above and is somewhat surprising considering that the proteins share 52% sequence identity. Assuming this statement to be correct, however, the difference may be explained by comparing the key residues involved in the formation of the octamer in Axe2, which as stated in §3.2, are Arg55, Trp95, Glu105, Asp145, Tyr184, Trp190 and Arg192. The analogous residues to Arg55, Trp95, Glu105, Asp145 and Arg192 are conserved in 3RJT, but Tyr184 and Trp190 are not. As stated, the Tyr184 and Trp190 play an important part in the contacts stabilizing the octameric assembly of Axe2. Since the corresponding Asn183 and Pro189 in 3RJT cannot form the same contacts as in the case of the Tyr and Trp side chains in the same position, this difference may explain, at least in part, why 3RJT may only form a dimer in solution and not the octameric assembly observed for Axe2. In light of this, the contacts formed by Tyr184 and Trp190 appear to be more critical than the tight intermolecular salt bridge formed by Arg55 and Glu105 in forming the octameric structure of Axe2.

It should be noted, however, that additional calculations by the *PISA* server using only the protein chains of 3RJT without any solvent suggest that an octameric assembly may also be possible for 3RJT. This possibility is in better agreement with the high percentage of sequence identity between the proteins. It is also compatible with the space group in which 3RJT crystallized, $P42_12$, since an octamer structure similar to Axe2 may be generated by applying crystallographic symmetry. Interestingly, when such a 3RJT octamer is constructed, a potential hydrogen bond is formed between the side-chain of Asn183 and the main-chain carbonyl group of an adjacent 3RJT monomer, stabilizing this octameric structure. Obviously, the oligomeric state of 3RJT requires further experimental evidence, as the current discussion on the oligomeric state of 3RJT is based mainly on the *PISA* calculations, which are somewhat questionable in this case.

3.6.2. Comparison with related oligomeric structures of hydrolases. A quick search among the reported structures of proteins confirms that the doughnut-shaped octameric structure of Axe2 is unusual. The structures of a few hexameric doughnut-shaped hydrolases have been reported in the past few years [*e.g.* PDB entries 1pz2 (Hövel *et al.*, 2003), 1ods (Vincent *et al.*, 2003), 2xlc (Montoro-García *et al.*, 2011), 3m81 (Levisson *et al.,* 2012), 3fcy (I. Krastanova, A. Cassetta, J. Wiegel & D. Lamba, unpublished work) and 1l7a (Midwest Center for Structural Genomics, unpublished work)], but only one other hydrolase with an octameric structure could be found (MsAcT; PDB entry 2q0q; Mathews *et al.*, 2007).

The structure of MsAcT, a hydrolase/acyltransferase from *Mycobacterium smegmatis*, resembles that of Axe2 in both its octameric assembly and its fold. The two enzymes share 20% sequence identity, belong to the GDSL family and possess the SGNH hydrolase fold for their eight monomer subunits. Thus, Axe2 and MsAcT have similar secondary-structure elements, both containing an inner layer of five β-sheets sandwiched by eight α-helices and additional 3_{10}-helices (five and two, respectively). Axe2 and MsAcT contain a similar number of amino acids per monomer (219 and 216, respectively) and an overlap between the monomer subunits of the two structures demonstrates significant similarity (Fig. 11*a*), with an r.m.s.d. value of 1.84 Å for 146 of 215 C^{α} atoms (calculated using the *DALI* server; Holm & Sander, 1995). Especially apparent are the similarities in the active sites, where the catalytic triad is situated in the same geometry in both proteins, with Ser adopting the innermost position and Asp the outermost. This is only to be expected considering that the two enzymes share the same GDSL fold and family.

Not surprisingly, however, Axe2 and MsAcT are quite different in respect to the number, location and orientation of their loops, although some similarities do exist. Noteworthy is the large and wide loop in MsAcT (residues 15–33) located near its active site (Fig. 11*a*) and the important contribution of a few residues positioned on it (Trp16, Ala23 and Pro24) to the formation of a hydrophobic channel found to aid binding of substrate during catalysis (Mathews *et al.*, 2007). Interestingly, the beginning and the end of this unique loop in MsAcT are located in similar positions and orientations to those of the 'double-curved' loop of Axe2 mentioned above (residues 19–36; Figs. 3*a*, 9*b* and 10*a*). Nevertheless, the overall conformations of the two loops differ significantly: it is oriented away from the active site in Axe2, while it is oriented towards the active site and practically 'covers' the substrate channel in MsAcT, demonstrating the high flexibility of this loop. These conformational differences also support our previous suggestions that this loop in Axe2 is involved in bringing the substrate to the active site of Axe2 or assisting its entrance into the substrate-binding site, similar to the role suggested for the corresponding loop in MsAcT.

The two enzymes are also similar in that both assume a cyclic octameric assembly (Fig. 11*b*). In both cases the assembly may be generally described by a central cavity formed by two layers of tetramers situated one on top of the other. However, in the case of Axe2 the two tetramer layers are 'staggered' with respect to one another, whereas in the case of MsAcT the two tetramer layers are 'eclipsed'.

A significant difference between the two octameric hydrolases can also be seen in the spatial arrangement of their active sites. While all eight catalytic sites of Axe2 are situated in the inner part of the octameric torus and face the central cavity (Fig. 4*a*), the catalytic sites of MsAcT are situated in the outer surface of the torus facing the solvent interface of the octamer. Interestingly, Axe2 is similar in these aspects to a number of hexameric doughnut-shaped hydrolases reported in the past, all of which contain the α/β-hydrolase fold instead of the SGNH fold (*e.g.* Vincent *et al.*, 2003; Montoro-García *et al.*, 2011; Levisson *et al.*, 2012). As in Axe2, these structures possess a double layer of staggered trimers and buried catalytic sites facing the central cavity. The dimensions of these oligomeric ring-like-shaped hydrolases are also of interest: they are all ~100 Å along their longest diameter but differ in the size of their cavities. This is depicted in Fig. 11(*b*), in which a comparison between the overall oligomeric structures of Axe2, MsAcT and a representative hexameric hydrolase (BpAXE; Montoro-García *et al.*, 2011) is shown. This comparison demonstrates that the cavity of Axe2 is

significantly larger compared with the cavities of the other proteins, being ∼35 Å as opposed to ∼20 Å in diameter.

The reasons for the significant differences in the cavity size and the orientation of the active sites presently remain unclear. In the general case, it was initially proposed that one of the roles of a ring-like structure is to hinder non-relevant substrates from entering the central cavity and gaining access to the active sites based on size, shape and chemical nature (Vincent *et al.*, 2003; Montoro-García *et al.*, 2011), as discussed above. However, as the cavity size of Axe2 is significantly larger than those of the other structures, and as its substrate is relatively small, questions arise regarding this general hypothesis. In addition, the fact that the catalytic sites of MsAcT are situated relatively close to the outer surface of the octamer (and do not face its central cavity) suggests that the central cavity may play a role, if any, other than being the means by which substrates reach the catalytic sites. These interesting functional arguments should be further examined by alternative experimental approaches.

The differences in the content and shape of the oligomeric assemblies of hydrolases present an even more general and complex issue. As mentioned above, the specific self-assembly of proteins to form higher order oligomers of different kinds, although quite common, is not as yet fully understood (Devenish & Gerrard, 2009; Griffin & Gerrard, 2012; Matthews & Sunde, 2012). Explanations for the advantages of oligomeric structures have included, over the years, arguments such as active-site formation across interfaces, allosteric regulation, oligomeric state as a regulatory mechanism, enhanced evolutionary sensitivity to mutation and increased stability. Nevertheless, no satisfactory unifying theory has yet been offered which explains the specific quaternary structure that has emerged for any individual protein. For example, in the superfamily of proteins that have the general $(\beta/\alpha)_8$-barrel fold, a series of studies revealed specific cases in which the apparent role of the homo-oligomeric structure is to complete the active site, to enhance thermal stability or to tune dynamic motion in order to optimize catalytic activity (Devenish & Gerrard, 2009). Similar and related studies should be applied to Axe2 and the multimeric proteins mentioned above in order to better understand the fine interplay between their function and their quaternary structure, if any. Some representative studies of this kind are briefly mentioned below.

3.7. Future validating experiments

As indicated in a number of places above, we are well aware of the fact that some of the structural and functional interpretations suggested in this report need to be validated by further experiments. This is especially critical for the functional oligomeric form of 3RJT in solution, which needs to be unequivocally determined by direct experimental techniques such as those presented here for Axe2. Other experiments should examine the specific interactions that lead to formation of the dimers, the tetramers and the octamer, as summarized in Table 3. In order to check the importance and the relative contribution of each of the listed contacts to formation of the relevant oligomer, a series of rational site-specific mutations should be performed and the resulting form of the enzyme should then be examined and characterized in terms of activity and specificity. A series of mutagenesis experiments of this kind are currently under way in our laboratory.

This work was 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. NEC thanks the UK Engineering and Physical Sciences Research Council (EPSRC), grant EP/G027005, for financial support. We thank the staff at the European Synchrotron Research Facility (ESRF, BM14 beamline) for their helpful support in the X-ray synchrotron data measurement and analysis. We thank David Azulai for his assistance in the refinement of the Axe2-S15A structure. We thank Ellina Kesselman for the negative-staining TEM images. We thank the staff at the X9 SAXS Workbench 2013 and the RapiData 2013 course (NSLS, Brookhaven National Laboratory, Upton, New York, USA) for their great help in the SAXS data collection and analysis. We also thank and acknowledge the helpful comments of the co-editor and the anonymous referees of this paper, especially the comments related to the relevant oligomeric form of 3RJT. YS holds the Erwin and Rosl Pollak Chair in Biotechnology at the Technion.

References

Akoh, C. C., Lee, G.-C., Liaw, Y.-C., Huang, T.-H. & Shaw, J.-F. (2004). *Prog. Lipid Res.* **43**, 534–552.
Alalouf, O., Balazs, Y., Volkinshtein, M., Grimple, Y., Shoham, G. & Shoham, Y. (2011). *J. Biol. Chem.* **286**, 41993–42001.
Allaire, M. & Yang, L. (2011). *J. Synchrotron Rad.* **18**, 41–44.
Aspinall, G. O. (1959). *Adv. Carbohydr. Chem.* **1**, 429–468.
Banacký, P. & Linder, B. (1981). *Biophys. Chem.* **13**, 223–231.
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.
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.
Biely, P. (2012). *Biotechnol. Adv.* **30**, 1575–1588.
Bitto, E., Bingman, C. A., McCoy, J. G., Allard, S. T. M., Wesenberg, G. E. & Phillips, G. N. (2005). *Acta Cryst.* D**61**, 1655–1661.
Bouveng, H. O. (1961). *Acta Chem. Scand.* **15**, 96–100.
Bravman, T., Mechaly, A. G., 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.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.

Czjzek, M., Bravman, T., Henrissat, B. & Shoham, Y. (2004). *Acta Cryst.* D**60**, 1461–1463.

Devenish, S. R. A. & Gerrard, J. A. (2009). *Org. Biomol. Chem.* **7**, 833–839.

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.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R. & Bateman, A. (2010). *Nucleic Acids Res.* **38**, D211–D222.

Gat, O., Lapidot, A., Alchanati, I., Regueros, C. & Shoham, Y. (1994). *Appl. Environ. Microbiol.* **60**, 1889–1896.

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.

Griffin, M. D. W. & Gerrard, J. A. (2012). *Adv. Exp. Med. Biol.* **747**, 74–90.

Guinier, A. (1939). *Ann. Phys.* **12**, 161.

Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.

Hövel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. & Schomburg, D. (2003). *EMBO J.* **22**, 4922–4932.

Khasin, A., Alchanati, I. & Shoham, Y. (1993). *Appl. Environ. Microbiol.* **59**, 1725–1730.

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.

Kozin, M. B. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 33–41.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

Lansky, S., Alalouf, O., Solomon, V., Alhassid, A., Govada, L., Chayan, N. E., Belrhali, H., Shoham, Y. & Shoham, G. (2013). *Acta Cryst.* F**69**, 430–434.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lapidot, A., Mechaly, A. & Shoham, Y. (1996). *J. Biotechnol.* **51**, 259–264.

Lee, Y.-L., Lee, L.-C. & Shaw, J.-F. (2012). *Biocatal. Agric. Biotechnol.* **1**, 95–104.

Levisson, M. *et al.* (2012). *Proteins*, **80**, 1545–1559.

Lindberg, B., Rosell, K.-G. & Svensson, S. (1973*a*). *Sven. Papperstidn.* **76**, 30–32.

Lindberg, B., Rosell, K.-G. & Svensson, S. (1973*b*). *Sven. Papperstidn.* **76**, 383–384.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Mathews, I., Soltis, M., Saldajeno, M., Ganshaw, G., Sala, R., Weyler, W., Cervin, M. A., Whited, G. & Bott, R. (2007). *Biochemistry*, **46**, 8969–8979.

Matthews, J. M. & Sunde, M. (2012). *Adv. Exp. Med. Biol.* **747**, 1–18.

Mølgaard, A., Kauppinen, S. & Larsen, S. (2000). *Structure*, **8**, 373–383.

Montoro-García, S., Gil-Ortiz, F., García-Carmona, F., Polo, L. M., Rubio, V. & Sánchez-Ferrer, Á. (2011). *Biochem. J.* **436**, 321–330.

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.

Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.

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.

Ramachandran, G. N., Sasisekharan, V. & Ramakrishnan, C. (1963). *J. Mol. Biol.* **7**, 95–99

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.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.

Shulami, S., Gat, O., Sonenshein, A. L. & Shoham, Y. (1999). *J. Bacteriol.* **181**, 3695–3704.

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.

Svergun, D. I. (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.

Vincent, F., Charnock, S. J., Verschueren, K. H. G., Turkenburg, J. P., Scott, D. J., Offen, W. A., Roberts, S., Pell, G., Gilbert, H. J., Davies, G. J. & Brannigan, J. A. (2003). *J. Mol. Biol.* **330**, 593–606.

Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.

Wei, Y., Schottel, J. L., Derewenda, U., Swenson, L., Patkar, S. & Derewenda, Z. S. (1995). *Nature Struct. Biol.* **2**, 218–223.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Yang, L. (2010). *SAXS/WAXS Data Collection at X9. SAXS Workbench.* Brookhaven National Laboratory, New York, USA.

Zaide, G., Shallom, D., Shulami, S., Zolotnitsky, G., Golan, G., Baasov, T., Shoham, G. & Shoham, Y. (2001). *Eur. J. Biochem.* **268**, 3006–3016.

Zhang, J., Siika-Aho, M., Tenkanen, M. & Viikari, L. (2011). *Biotechnol. Biofuels*, **4**, 60.

Zolotnitsky, G., Cogan, U., Adir, N., Solomon, V., Shoham, G. & Shoham, Y. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 11275–11280.