BIOLOGICAL CRYSTALLOGRAPHY
ISSN 1399-0047

Structure–function relationships in Gan42B, an intracellular GH42 β-galactosidase from *Geobacillus stearothermophilus*

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

[a]Institute of Chemistry and the Laboratory for Structural Chemistry and Biology, The Hebrew University of Jerusalem, Jerusalem 91904, Israel, and [b]Department of Biotechnology and Food Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel. *Correspondence e-mail: yshoham@tx.technion.ac.il, gil2@vms.huji.ac.il

Received 27 August 2015
Accepted 5 October 2015

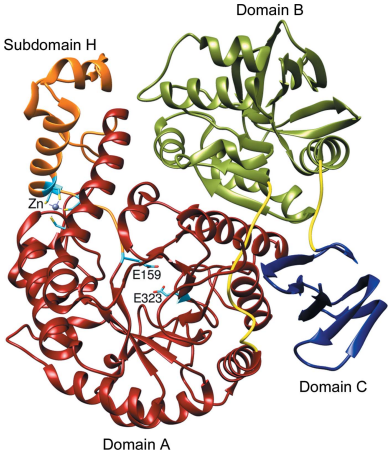
Edited by Z. Dauter, Argonne National Laboratory, USA

Keywords: *Geobacillus stearothermophilus*; galactan utilization; galactosaccharides; galactose; β-galactosidase; glycoside hydrolase; GH42; homotrimer; catalytic mutant; glycosynthase.

PDB references: Gan42B-WT, 4oif; Gan42B-E323A, 5dfa

Supporting information: this article has supporting information at journals.iucr.org/d

Geobacillus stearothermophilus T-6 is a Gram-positive thermophilic soil bacterium that contains a battery of degrading enzymes for the utilization of plant cell-wall polysaccharides, including xylan, arabinan and galactan. A 9.4 kb gene cluster has recently been characterized in *G. stearothermophilus* that encodes a number of galactan-utilization elements. A key enzyme of this degradation system is Gan42B, an intracellular GH42 β-galactosidase capable of hydrolyzing short β-1,4-galactosaccharides into galactose units, making it of high potential for various biotechnological applications. The Gan42B monomer is made up of 686 amino acids, and based on sequence homology it was suggested that Glu323 is the catalytic nucleophile and Glu159 is the catalytic acid/base. In the current study, the detailed three-dimensional structure of wild-type Gan42B (at 2.45 Å resolution) and its catalytic mutant E323A (at 2.50 Å resolution), as determined by X-ray crystallography, are reported. These structures demonstrate that the three-dimensional structure of the Gan42B monomer generally correlates with the overall fold observed for GH42 proteins, consisting of three main domains: an N-terminal TIM-barrel domain, a smaller mixed α/β domain, and the smallest all-β domain at the C-terminus. The two catalytic residues are located in the TIM-barrel domain in a pocket-like active site such that their carboxylic functional groups are about 5.3 Å from each other, consistent with a retaining mechanism. The crystal structure demonstrates that Gan42B is a homotrimer, resembling a flowerpot in general shape, in which each monomer interacts with the other two to form a cone-shaped tunnel cavity in the centre. The cavity is ∼35 Å at the wide opening and ∼5 Å at the small opening and ∼40 Å in length. The active sites are situated at the interfaces between the monomers, so that every two neighbouring monomers participate in the formation of each of the three active sites of the trimer. They are located near the small opening of the cone tunnel, all facing the centre of the cavity. The biological relevance of this trimeric structure is supported by independent results obtained from gel-permeation chromatography. These data and their comparison to the structural data of related GH42 enzymes are used for a more general discussion concerning structure–activity aspects in this GH family.



1. Introduction

β-Galactosidases (EC 3.2.1.23) form a large group of enzymes that hydrolyze the β-glycosidic bond between terminal non-reducing D-galactose residues and other organic molecules. Based on their sequence similarities, β-galactosidases are classified into four main glycoside hydrolase (GH) families, GH1, GH2, GH35 and GH42, according to the CAZy database (Henrissat & Davies, 1997). These families have distinctly different sequences and also exhibit different substrate specificities. In general, GH1 and GH2 β-galactosidases are

found in mesophiles and demonstrate lactase activity, while enzymes belonging to families GH35 and GH42 are found in thermophiles, preferably degrading β-1,4 linkages between two galactose moieties with very low or a complete absence of lactase activity (Ohtsu *et al.*, 2014; Yang *et al.*, 2003; Hinz *et al.*, 2004; Shipkowski & Brenchley, 2006; Schwab *et al.*, 2010). Since the natural habitats of thermophilic microorganisms are usually relatively poor in lactose, the native substrates of their GH42 enzymes are thought to be those galacto-oligomers derived from the high-molecular-weight polymers composing the plant cell wall. This logical hypothesis is supported by the observation that GH42 β-galactosidases are located in proximity to GH53 galactanases in many genetic systems, in operons dedicated to the utilization of pectic type I galactan and arabinogalactan, two important polysaccharide components of the plant cell wall (Daniel *et al.*, 1997; Shipkowski & Brenchley, 2006; Delangle *et al.*, 2007; O'Connell Motherway *et al.*, 2011).

Geobacillus stearothermophilus T-6 is a Gram-positive thermophilic soil bacterium that harbours a well regulated system for the utilization of plant cell-wall polysaccharides, including xylan, arabinan and galactan (Shulami *et al.*, 1999, 2011; Tabachnikov & Shoham, 2013). The bacterium utilizes a limited number of endo-acting extracellular enzymes that break down the high-molecular-weight polysaccharides into decorated oligosaccharides. These oligosaccharides enter the cell *via* specialized ABC transporters (Rees *et al.*, 2009) and are further hydrolyzed into sugar monomers by a set of intracellular glycoside hydrolases. For example, for the complete consumption of xylan the bacterium secretes an extracellular xylanase (XT6; Gat *et al.*, 1994; Teplitsky *et al.*, 1997, 2004; Bar *et al.*, 2004), which partially degrades the long xylan polymer to shorter decorated xylo-oligomers, which are then imported into the cell *via* the oligoxylose ABC transport system (Shulami *et al.*, 2007). Inside the cell, the decorated xylo-oligomers are hydrolyzed by a battery of side-chain-cleaving enzymes, including arabinofuranosidases (Shallom, Belakhov, Solomon, Gilead-Gropper *et al.*, 2002; Shallom, Belakhov, Solomon, Shoham *et al.*, 2002; Hövel, Shallom, Niefind, Baasov *et al.*, 2003; Hövel, Shallom, Niefind, Belakhov *et al.*, 2003; Lansky, Salama, Dan *et al.*, 2014), glucuronidases (Teplitsky *et al.*, 1999; Zaide *et al.*, 2001; Golan, Shallom *et al.*, 2004; Shallom *et al.*, 2004) and acetyl-esterases (Alalouf *et al.*, 2011; Lansky, Alalouf, Solomon *et al.*, 2013, 2014; Lansky, Alalouf, Salama *et al.*, 2014), and finally by an intracellular xylanase (IXT6; Teplitsky *et al.*, 2000; Solomon *et al.*, 2007) and a number of xylosidases (Bravman *et al.*, 2001, 2003; Shallom *et al.*, 2005; Brüx *et al.*, 2006; Ben-David *et al.*, 2007, 2008; Dann *et al.*, 2014), resulting in xylose monomers.

We have recently characterized a 9.4 kb gene cluster in *G. stearothermophilus*, *ganREFGBA*, which encodes a number of galactan-utilization elements (Tabachnikov & Shoham, 2013). In this system, the *ganEFG* genes encode an ATP-binding cassette sugar-transport system, the sugar-binding protein of which, GanE, was shown to bind galacto-oligosaccharides. The *ganA* gene encodes the Gan53A protein (88.7 kDa), an extracellular glycoside hydrolase family 53 (GH53) β-1,4-galactanase. Gan53A is active mainly on high-molecular-weight galactan and produces galactotetraose as the main product. Similarly, the *ganB* gene encodes the Gan42B protein (80.2 kDa), a glycoside hydrolase family 42 (GH42) β-galactosidase which was suggested to function mainly in hydrolyzing short β-1,4-galactosaccharides into galactose monomers. Detailed biochemical characterization of the recombinant Gan42B enzyme has supported this role, demonstrating significant hydrolytic activity on galactobiose and larger galacto-oligomers, with no detectable activity towards lactose. The application of both Gan53A and Gan42B on galactan resulted in the full degradation of the polymer into galactose units, to be metabolized later in the bacterium into UDP-glucose *via* the Leloir pathway by the *galKET* gene products (Tabachnikov, 2012; Tabachnikov & Shoham, 2013).

As a key player in the biochemical hydrolysis of polymeric β-1,4-galactosaccharides into galactose monomers, the Gan42B β-galactosidase and related enzymes play an important part in the hemicellulolytic utilization system of many microorganisms that use plant biomass for growth. The interest in the detailed biochemical characterization and structural analysis of these enzymes stems therefore not only from basic scientific interest, but also from their numerous potential biotechnological applications. The Gan42B protein monomer is made up of 686 amino acids, and based on sequence homology it was suggested that Glu323 is its catalytic nucleophile and Glu159 is its catalytic acid/base. Considering these data, we recently designed, produced and tested the Gan42B nucleophile mutant E323A (Gan42B-E323A), which showed preliminary glycosynthase activity, allowing the reverse enzymatic reaction synthesizing galacto-oligomers from galactose (Tabachnikov, 2012). The structure–function analysis of wild-type Gan42B (Gan42B-WT) and Gan42B-E323A was accordingly the main target of the research described here.

Most of the current knowledge on the mechanism and mode of action of β-galactosidases in general, and of GH42 β-galactosidases in particular, is based on the corresponding three-dimensional structures. Although a pioneering cryo-EM structural analysis has recently been reported for a β-galactosidase at a relatively high resolution (Bartesaghi *et al.*, 2015), the most relevant three-dimensional structures of GH42 β-galactosidases have been determined by 'traditional' single-crystal X-ray crystallography. Three such structures have been reported to date: A4-β-gal from *Thermus thermophilus* A4 (33% identity to Gan42B; PDB entry 1kwk; Hidaka *et al.*, 2002), Bca-β-gal from *Bacillus circulans* sp. *alkalophilus* (39% identity; PDB entry 3tty; Maksimainen *et al.*, 2012) and *Bl*Gal42A from *Bifidobacterium animalis* subsp. *lactis* Bl-04 (32% identity; PDB entry 4uni; Viborg *et al.*, 2014). In the present report, we describe the full crystallographic structural characterization of Gan42B as the fourth structure of this group of enzymes, focusing on the detailed three-dimensional structures of Gan42B-WT and Gan42B-E323A at 2.45 and 2.50 Å resolution, respectively. These structures are then discussed in respect to the function, specificity and catalytic mechanism of Gan42B and are used as the basis for structure–

function interpretations and comparisons with related GH42 β-galactosidases.

2. Experimental

2.1. Overexpression and purification of Gan42B-WT and Gan42B-E323A

The *ganB* gene with a tag of six histidine residues fused at its N-terminus was cloned into the expression vector pET-9d as described previously (Tabachnikov & Shoham, 2013; Solomon *et al.*, 2013). Briefly, site-directed mutagenesis of *ganB* was performed using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, California, USA). The cloned His-tagged proteins were overproduced in *Escherichia coli* BL21(λDE3) using the T7 polymerase expression system and purified in a single step using nickel-affinity chromatography, as previously performed for other His-tagged proteins from *G. stearothermophilus* (Shulami *et al.*, 2011). For both Gan42B-WT and Gan42B-E323A the final protein solution obtained was more than 95% pure based on SDS–PAGE, with a total yield of about 500 mg protein per litre of overnight culture. No attempts were made to remove the His-tag peptide after purification, and both proteins were used as such in all subsequent experiments, including crystallization. Gel-permeation chromatography was performed on a Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences) running at 0.5 ml min^{-1} at room temperature. The resulting chromatogram presented a relatively sharp main peak (Fig. 1), indicating that the native protein is highly homogeneous with a molecular weight of about 235 kDa, and suggesting that Gan42B is a homotrimer in solution under physiological conditions.

2.2. Crystallization and X-ray diffraction data collection

As described previously in greater detail (Solomon *et al.*, 2013), suitable crystals of the purified Gan42B-WT and Gan42B-E323A proteins (including their original His tags) were obtained by a series of hanging-drop vapour-diffusion experiments, optimizing parameters such as the pH, ionic strength, protein concentration and temperature (Almog *et al.*, 1993, 1994; Gilboa *et al.*, 1998; Golan, Zharkov *et al.*, 2004; Reiland *et al.*, 2004), as well as precipitating additives (Lansky, Salama, Solomon *et al.*, 2013; Lansky, Zehavi *et al.*, 2014). For both proteins, the best diffracting crystals were obtained at 21°C in 4 µl drops consisting of a mixture of 2 µl protein solution (18–22 mg ml^{-1} in 100 m*M* NaCl, 0.02% sodium azide, 50 m*M* Tris buffer pH 7.0) and 2 µl crystallization reservoir solution (16–18% PEG 3350, 250 m*M* NaCl, 100 m*M* HEPES buffer pH 7.0). These crystals usually grew (in 5–15 d) as a cluster of 5–10 thin plates, each with typical dimensions of about 0.3 × 0.2 × 0.02 mm. The crystallization conditions and final sizes and shapes of the resulting crystals were closely similar for Gan42B-WT and Gan42B-E323A. For unknown reasons both crystals were very hard to reproduce, and usually only a small percentage of the hanging drops resulted in crystallographically useful crystals (Solomon *et al.*, 2013).

In all cases it was necessary to separate the target crystal from the multi-crystal cluster obtained in order to submit it to full crystallographic analysis. This was performed by mechanical cutting of the cluster into separate plates, as close as possible to the cluster-connecting part of the plate, and with special efforts to apply minimal deformation force to the cluster and the separated crystals (Solomon *et al.*, 2013). The crystal flash-cooling procedure included a short 20–30 s pre-soaking of the selected crystal in a cryoprotecting solution, consisting of 87% of the original crystallization reservoir solution and 13%(*v*/*v*) glycerol, prior to its quick immersion in liquid nitrogen.

One of the Gan42B-WT crystals was separated and cooled in this way, and was used for a full X-ray diffraction data measurement at 2.45 Å resolution. This data set was collected at 100 K, using X-ray synchrotron radiation (λ = 0.9334 Å) and a CCD area detector (ADSC Q210), on the ID14-1 beamline at the ESRF, France. The raw CCD oscillation data set (Δφ = 0.5°, 10 s exposure, 360 frames, 180°) was processed with *iMosflm* (Powell *et al.*, 2013) and *HKL*-2000 (Otwinowski & Minor, 1997), demonstrating that the crystals belonged to the primitive orthorhombic space group $P2_12_12_1$, with average crystallographic unit-cell parameters a = 71.84, b = 181.35, c = 196.57 Å. A total of 531 503 accepted reflections [$F > 0\sigma(F)$] were measured in the 49.0–2.45 Å resolution range, and resulted in 91 957 independent reflections, with 96.5% completeness to 2.45 Å resolution and 88.6% completeness for the highest resolution shell (2.48–2.45 Å). The overall redundancy in the data set was 5.8, the overall mosaicity was 0.71°, the average $I/\sigma(I)$ was 11.7, and the final $R_{\rm sym}$ (or $R_{\rm merge}$) for the whole data was 11.3% (34.8% for the highest resolution shell) (Table 1).

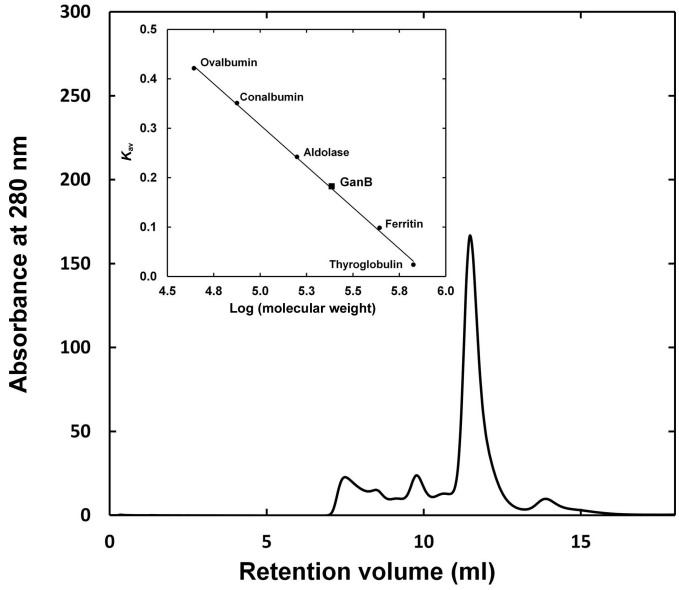


Figure 1
Gel-permeation chromatogram for Gan42B. Inset, selectivity curve for a Superdex 200 Increase 10/300 GL column. The curve presents the partition coefficient $K_{\rm av}$ against the log of the molecular weight for a set of standard proteins. The calculated molecular weight of Gan42B based on the selectivity curve is 234.4 kDa, suggesting a trimeric oligomer state in solution.

Table 1
Representative parameters from the crystallographic data measurements.

Values in parentheses are for the last resolution shell.

	Gan42B-WT	Gan42B-E323A
Beamline	ID14-1, ESRF	I04, Diamond
Wavelength (Å)	0.9334	0.9763
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 71.84, b = 181.35, c = 196.57	a = 71.73, b = 181.31, c = 197.66
Resolution (Å)	48.9–2.45 (2.58–2.45)	71.7–2.50 (2.56–2.50)
No. of reflections		
Total	531503	285572
Unique	91957	87146
Multiplicity	5.8 (4.9)	3.3 (3.0)
$\langle I\rangle/\langle\sigma(I)\rangle$	11.7 (4.3)	13.9 (5.2)
Mosaicity (°)	0.71	0.58
Completeness (%)	96.5 (88.6)	97.4 (96.6)
$R_{\rm merge}$† (%)	11.3 (34.8)	6.1 (19.8)
$R_{\rm meas}$‡ (%)	13.5 (41.5)	7.2 (23.9)
$R_{\rm p.i.m.}$§ (%)	5.4 (17.6)	3.7 (13.1)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. § $R_{\rm p.i.m.} = \sum_{hkl}\{1/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is an individual intensity measurement and $\langle I(hkl)\rangle$ is the average of symmetry-related observations of a unique reflection.

In the case of Gan42B-E323A, crystals also appeared in clusters and were confirmed to be highly isomorphous to the corresponding Gan42B-WT crystals (space group $P2_12_12_1$; average unit-cell parameters a = 71.73, b = 181.31, c = 197.66 Å). One of these Gan42B-E323A crystals was used for the measurement of a complete oscillation data set to 2.50 Å resolution, collected at 100 K, using X-ray synchrotron radiation (λ = 0.9763 Å) and an ADSC Q315 area detector, on the I04 beamline at Diamond Light Source, England. Data were collected with an oscillation range of 90°, a $\Delta\varphi$ of 0.5°, a time exposure of 0.5 s and a crystal-to-detector distance of 378.5 mm. The raw CCD diffraction images were processed with *iMosflm* (Powell *et al.*, 2013) and *HKL*-2000 (Otwinowski & Minor, 1997). A total of 285 572 accepted reflections [$F > 0\sigma(F)$] were measured in the 71.70–2.50 Å resolution range, resulting in 87146 independent reflections with 97.4% completeness to 2.50 Å resolution (Table 1). The overall redundancy in the data set was 3.3, the overall mosaicity was 0.58, the average $I/\sigma(I)$ was 13.9 and the final $R_{\rm sym}$ was 6.1% (19.8% for the highest resolution shell). Apparently, the significantly lower $R_{\rm sym}$ in this case was not owing to lower redundancy compared with the Gan42B-WT data, as the calculation of redundancy-independent R values ($R_{\rm meas}$ and $R_{\rm p.i.m.}$) resulted in roughly the same R ratios (Table 1).

2.3. Structure determination of Gan42B-E323A

At the time of the structure determination of Gan42B, the structure of only one other GH42 β-galactosidase was reported and available: that of a GH42 β-galactosidase from the extreme thermophile *T. thermophilus* A4 (A4-β-Gal, 33% sequence identity; Hidaka *et al.*, 2002). The coordinates of this structure (1.6 Å resolution; PDB entry 1kwg) were used as the initial model for molecular-replacement (MR) crystallographic calculations, eventually leading to the full structure determination of Gan42B. Initially, since the diffraction data for the Gan42B-E323A mutant appeared to be significantly better than those for Gan42B-WT, it was decided to start with these data for structure determination. Only at a later point was a higher resolution and improved data set collected for Gan42B-WT (Table 1).

MR calculations on the 2.5 Å resolution data for Gan42B-E323A (using *Phaser*; McCoy *et al.*, 2007; Cowtan *et al.*, 2011) were performed with the TIM-barrel domain (residues 1–480) of the A4-β-Gal structure (see below). Self-rotation function calculations indicated the presence of a threefold noncrystallographic rotational symmetry, suggesting that three Gan42B monomers were present in the crystallographic asymmetric unit, as corroborated by Matthews coefficient calculations (Solomon *et al.*, 2013). MR calculations were performed in three steps, searching for one monomer at a time, while fixing those of the previously located monomers. As expected, this stepwise procedure resulted in three Gan42B monomers in the crystallographic asymmetric unit (chains *A*, *B* and *C*), corresponding to a log-likelihood gain (LLG) value of 303.2 and a TFZ score of 13.7 (McCoy *et al.*, 2007). The structure of the Gan42B-E323A mutant was initially built by the *AutoBuild* utility of *PHENIX* (Terwilliger *et al.*, 2008) and was subsequently completed manually with *Coot* (Emsley & Cowtan, 2004).

Structural refinement of the Gan42B-E323A mutant was performed with *REFMAC*5 (Murshudov *et al.*, 2011) and *phenix.refine* (Afonine *et al.*, 2012) as implemented in the *CCP*4 (Cowtan *et al.*, 2011) and *PHENIX* (Adams *et al.*, 2010) software packages, respectively. Each round of refinement was followed by manual fitting and rebuilding with *Coot* (Emsley & Cowtan, 2004). A negative electron density around the side chain of residue 323 confirmed the point mutation of this residue from Glu to Ala. A strong electron density (>6σ in $F_{\rm o} - F_{\rm c}$ maps) was clearly visible at the expected Zn^{2+} site and was therefore modelled as such. In the final stages of refinement, water molecules were included in the Gan42B-E323A model based on suitable distances and orientations to form hydrogen bonds. The final Gan42B-E323A structure (a full protein trimer in the asymmetric unit) consisted of 1120 water molecules, three Zn atoms and nine glycerol molecules (probably originating from the cryoprotecting solution). 36 residues were modelled in alternate conformations. Almost all protein residues were modelled in the final model, except for some flexible regions that were not modelled owing to insufficiently clear electron density. These regions included the first 2–4 residues at the N-terminus and residues 666–670 at the C-terminus in all three chains, and residues 657–660 in chain *C* only. Refinement of the model converged to a final R factor of 14.15% and an $R_{\rm free}$ of 21.48% (for 5% of the data). Representative parameters for the refinement and final model are listed in Table 2.

PROCHECK (Laskowski *et al.*, 1993) was used for general validation of the main structural parameters of Gan42B-E323A. Such analysis of the polypeptide-chain conformations (Ramachandran *et al.*, 1963) confirmed the conformational validity of the model, with 97.2% of the residues in the most

Table 2
Refinement parameters of the final Gan42B-WT and Gan42B-E323A models.

	Gan42B-WT	Gan42B-E323A
Resolution (last shell) (Å)	48.86–2.45 (2.48–2.45)	67.43–2.50 (2.52–2.50)
R_{work} (%)	13.29	14.15
R_{free} (%)	20.45	21.48
Ramachandran plot (%)		
Favoured	96.66	97.07
Allowed	3.30	2.83
Outliers	0.05	0.10
No. of water molecules	1553	1134
R.m.s.d., angles (°)	1.09	1.12
R.m.s.d., bond lengths (Å)	0.008	0.009
Average B factor (Å^2)		
Overall	14.4	28.1
Protein	14.0	28.0
Ligand	24.0	41.4
Water	17.9	29.3
PDB code	4oif	5dfa

favoured region of the Ramachandran plot, 2.8% of the residues in the additionally allowed region and only 0.1% in the outliers region. These results are reasonable for a structure of this size refined at ∼2.5 Å resolution. The protein showed good bond-length and bond-angle values (Engh & Huber, 1991), with root-mean-square deviations (r.m.s.d.s) of 0.009 Å for bond lengths and 1.12° for bond angles. The average B factor for the whole structure was 28.1 Å^2 (Table 2). These overall structural parameters are quite good for a protein of this complexity, confirming that the final model is reliable and is suitable for structure–function analysis and structure comparisons.

2.4. Structure determination and refinement of Gan42B-WT

The structure of Gan42B-WT was solved on the basis of the three-dimensional structure of the Gan42B-E323A mutant using *REFMAC*5 (Murshudov *et al.*, 2011), *phenix.refine* (Afonine *et al.*, 2012) and *Coot* (Emsley & Cowtan, 2004) for the refinement processes, similarly to the procedures described above. Here again the first 2–3 residues at the N-terminus were not modelled owing to insufficient electron density, and similarly disordered electron density was also observed around residues Thr261 and His262. 37 residues were modelled in alternate conformations. The final model of Gan42B-WT (as a full trimer) consisted of three Zn atoms, 25 glycerol molecules and 1504 water molecules. Refinement of the Gan42B-WT model converged to a final R factor of 13.29% and a final R_{free} of 20.45% (for 5% of the data). Representative parameters for the refinement and the final model are listed in Table 2.

The validation parameters for the Gan42B-WT structure using *PROCHECK* (Laskowski *et al.*, 1993) appeared to be similar to those for the Gan42B-E323A structure. The Ramachandran plot indicated that 96.7% of the residues are in the most favoured region, 3.2% in the additionally allowed region and only 0.05% in the outliers region. The r.m.s.d. values for the bond lengths and bond angles were 0.008 Å and 1.09°, respectively, and the overall B factor for all atoms was 14.4 Å^2 (Table 2). These values conform well to the usual stereochemical parameters observed for protein structures of similar size determined in the 2.5 Å resolution range, confirming that the structure can be used for meaningful structural interpretations.

2.5. ICP measurements

The concentrations of a series of potential metal ions (Al, Ca, Cd, Co, Cu, Fe, Mg, Mn, Ni, Pb and Zn) were determined by inductively coupled plasma optical emission spectrometry (ICP-OES) using an iCAP 6000 spectrometer (Thermo Scientific). These measurements were performed on the native protein, as well as on EDTA-treated protein. EDTA treatment was performed in order to test which of the metal ions can be removed *via* complexation with EDTA, if any. Enzyme samples (0.3–3 mg ml^{-1}) in 50 m*M* Tris–HCl pH 7.5 were treated with 10 m*M* EDTA for 3 h at room temperature. It was then dialyzed overnight against 500 ml 50 m*M* Tris–HCl pH 7.5 buffer at room temperature with two buffer changes, one using 2 m*M* EDTA and the second using 1 m*M* EDTA. EDTA was removed by dialysis against the same buffer without EDTA but containing Chelex 100 (25 g l^{-1}, Na$^+$ form, 100–200 mesh, Bio-Rad). Additional samples were prepared without EDTA treatment and dialyzed against 50 m*M* Tris–HCl pH 7.5. Samples of the dialysis buffer were also analyzed for metal content, and the observed concentrations were subtracted from those obtained in the enzyme samples.

2.6. Kinetic measurements

The activity of Gan42B-WT and mutants on the synthetic substrate 4-nitrophenyl-β-D-galactopyranoside (*p*NPG, Sigma–Aldrich) was determined by measuring the release of 4-nitrophenol at 420 nm using a continuous colorimetric assay. Enzymatic reactions were performed with 8 m*M* *p*NPG at 40°C in 100 m*M* citric acid–Na_2HPO_4 buffer pH 6.0 solution containing 0.3 mg ml^{-1} BSA. The molar extinction coefficient of 4-nitrophenol determined under the described conditions was $\Delta\varepsilon$ = 1.76 mM^{-1} cm^{-1}. The initial reaction rates were measured for the first 5–10 min.

2.7. Differential scanning calorimetry (DSC)

The melting temperatures of Gan42B-WT and the C164A/C166A mutant were determined using a Microcal VP-DSC differential scanning calorimeter (MicroCal Inc., Northampton, Massachusetts, USA). Protein samples (0.3–0.5 mg ml^{-1}) were dialyzed extensively overnight against buffer (50 m*M* Tris–HCl pH 7.0, 100 m*M* NaCl, 0.02% NaN_3), and the actual dialysis buffer was used as the reference solution for the DSC scan. Samples were analyzed at a heating rate of 1°C min^{-1} over the temperature range 35–90°C. A scan with a buffer solution in both cells was subtracted from each data set, and the molar heat capacity, C_p, was calculated accordingly. The melting temperature was defined as the temperature at the maximum molar heat capacity.

2.8. Calculations and figure preparation

The matrices for the superposition of the protein structures were calculated by a least-squares distance-minimization algorithm (*LSQ*) implemented within *Coot* (Emsley & Cowtan, 2004), using the active-site C^{α} atoms as the guide coordinates. Figs. 1 and 2(*b*) were prepared with *Microsoft PowerPoint* 2007, Figs. 2(*a*), 2(*c*), 3(*a*), 3(*d*), 4 and 6 with *Chimera* (Pettersen *et al.*, 2004), Figs. 3(*b*) and 3(*c*) with *PyMOL* (DeLano, 2002), Fig. 5 with *ESPript* (Robert & Gouet, 2014) and Fig. 7 with *Origin* 5.0 (MicroCal).

2.9. PDB accession codes

Atomic coordinates for the Gan42B-WT and Gan42B-E323A structures have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (Rose *et al.*, 2013) with accession codes 4oif and 5dfa, respectively.

3. Results and discussion

3.1. The overall structure of the Gan42B monomer

The three-dimensional structures of Gan42B-WT and its catalytic mutant Gan42B-E323A are presented here at 2.45 and 2.50 Å resolution, respectively. These structures were determined *via* the molecular-replacement methodology, using the structure of the β-galactosidase from *T. thermophilus* A4 (A4-β-gal; 33% identity to Gan42B; PDB entry 1kwk; Hidaka *et al.*, 2002) as a search model. The two crystal structures proved to be isostructural, with three independent protein monomers (labelled as chains *A*, *B* and *C*) in the asymmetric unit of a $P2_12_12_1$ unit cell. In both cases, structural refinement of the protein models converged to relatively good *R* factors and geometrical values (Table 2), permitting a meaningful and reliable analysis of the final structures obtained. The overall structure of the resulting Gan42B-WT monomer is shown in Fig. 2, demonstrating three distinct domains: domain A (residues 1–401), domain B (414–615) and domain C (621–686). Domain A is connected to domain B by a

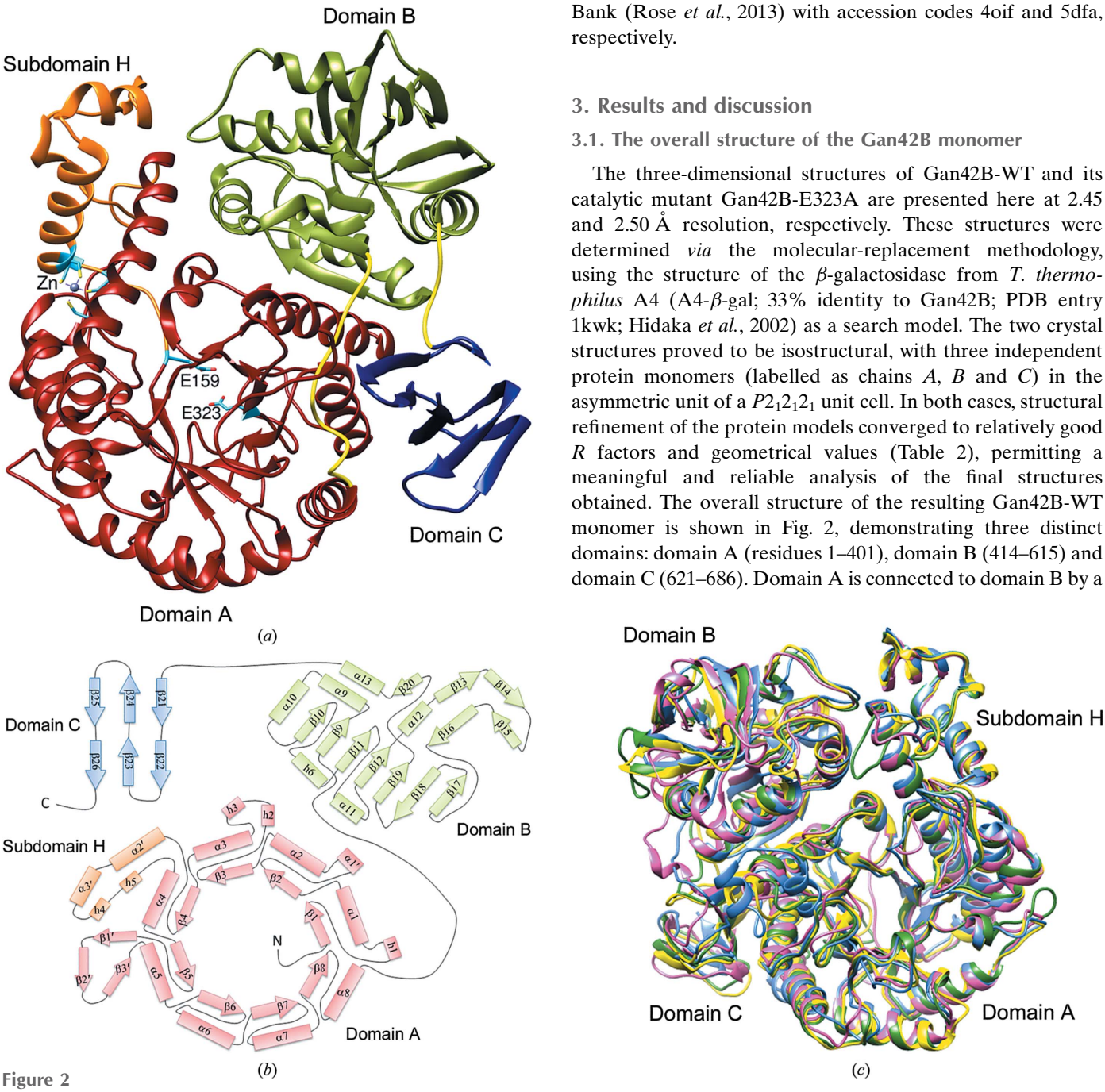


Figure 2
The three-dimensional structure of the Gan42B monomer. (*a*) Ribbon representation of the monomer structure. The protein is built of three distinct domains: domain A (shown in red and orange), domain B (green) and domain C (blue). The typical TIM-barrel structure of domain A is in red and the small subdomain H is in orange. The catalytic residues of Gan42B, Glu159 and Glu323, and the four Cys residues binding the Zn atom, are marked in light blue. The two loops connecting the domains are shown in yellow. (*b*) A schematic topology diagram of the Gan42B monomer. The secondary-structure elements are presented and numbered, with α-helices (α_n) and 3_{10}-helices (h_n) as rectangles, β-strands (β_n) as arrows and loops as curved single lines. (*c*) Superposition of the Gan42B monomer structure (yellow) onto the structures of A4-β-Gal (blue; PDB entry 1kwk), Bca-β-gal (green; PDB entry 3tty) and *Bl*Gal42A (pink; PDB entry 4uni), demonstrating the relatively high similarity between these structures, especially in domain A.

long random coil (residues 402–413), while domain B is connected to domain C by a relatively short loop (residues 616–620) (Fig. 2*a*).

Domain A, the largest of the three domains, is also the catalytic domain of the enzyme. It has the canonical $(\beta/\alpha)_8$-barrel (or TIM-barrel) fold commonly observed in glycoside hydrolases (GH) belonging to the GH-A clan (Durand *et al.*, 1997). In this fold eight alternating α-helices and β-strands form a round cylinder with a small cavity in the centre. The catalytic residues of Gan42B, Glu159 (acid/base) and Glu323 (nucleophile), are located around this central cavity, positioned at the C-terminal ends of the $\beta4$ and $\beta7$ strands of the TIM-barrel, respectively. Domain A contains an extra polypeptide outside the TIM-barrel fold, referred to as subdomain H (residues 160–217; Hidaka *et al.*, 2002), consisting of a cluster of two α-helices and two 3_{10}-helices inserted between $\beta4$ and $\alpha4$ of the TIM barrel (Fig. 2). This subdomain appears to play a role in the oligomerization of the protein, as well as in catalysis, as further described below (§§3.2 and 3.3). Interestingly, an extra subdomain, whenever relevant for TIM-barrel proteins, is usually inserted between the $\beta4$ and $\alpha4$ secondary elements and is often involved (directly or indirectly) in substrate specificity and/or substrate orientation (see, for example, Teplitsky *et al.*, 2004), as discussed further below.

Domain A also contains a metal-binding site formed by four highly conserved cysteine residues: Cys124, Cys164, Cys166 and Cys169. Cys124 is located on the loop right before the $\alpha3$ helix of the TIM barrel, while Cys164, Cys166 and Cys169 are located on the $\alpha2'$ helix and the loop that form part of subdomain H (Fig. 2). The metal in this site, assigned as a zinc cation, seems to function as a stabilizing element for the structure and conformation of subdomain H, as further discussed in §3.4.

Domain B is composed of 13 β-strands, five α-helices and one 3_{10}-helix (Fig. 2). The β-strands are located at the centre of the domain, surrounded by the α-helices. The central β-strands form a β-sheet composed of seven β-strands, which tapers off to form a small β-barrel. A similar fold to domain B is also present in one of the domains of the pyridoxal 5′-phosphate synthase from *Thermotoga maritima* (YaaE; PDB entry 2iss), which catalyzes the production of ammonia from hydrolysis of glutamine (Zein *et al.*, 2006). The r.m.s.d. between the corresponding domains of Gan42B and YaaE is 2.8 Å (based on 158 aligned C^{α} atoms) and their *DALI* *Z*-score is 13.4 (Holm & Rosenström, 2010), indicating relatively good structural homology between these two domains. Interestingly, this domain in YaaE contains the catalytic triad of the enzyme (Glu172–His170–Cys78), while in Gan42B these catalytic residues are absent. Domain B in Gan42B seems rather to play a role in the oligomerization of the protein and not in catalysis, as further discussed below. The relatively high structural similarities between the corresponding domains of these proteins, despite their different functional roles, may suggest a possible common evolutionary origin of these two domains.

Domain C is the smallest domain of the protein, consisting of six β-strands. These β-strands form two small β-sheet layers, each formed by three antiparallel β-strands (Fig. 2*b*). Domains with similar topology have been identified in enzymes from the GH13 family, such as the neopullulanase from *Bacillus stearothermophilus* (PDB entry 1j0h; Hondoh *et al.*, 2003) and the sucrose hydrolase from *Xanthomonas campestris* pv. *campestris* (PDB entry 2wpg; Champion *et al.*, 2009). The corresponding domains of these two enzymes are also located at the C-termini of the proteins and consist of eight (neopullulanase) and six (sucrose hydrolase) antiparallel β-strands. The r.m.s.d. values between domain C of Gan42B and these two proteins are 2.4 and 2.3 Å (for 63 and 60 common C^{α} atoms), and the corresponding *DALI* *Z*-scores are 7.0 and 6.9 (Holm & Rosenström, 2010), respectively. Interestingly, in the sucrose hydrolase from *X. campestris* pv. *campestris* the position and orientation of this domain relative to its catalytic $(\beta/\alpha)_8$ domain are similar to those observed in Gan42B. In both enzymes, however, this domain does not seem to play a role either in catalysis or in oligomeric structure formation, and its function in Gan42B and its homologous proteins is currently unknown (Hidaka *et al.*, 2002).

Comparison between the structures of Gan42B-WT and the Gan42B-E323A nucleophile mutant shows no significant differences between the two proteins, except for the obvious local differences around the mutated residue 323. A superposition of the two structures results in a relatively small r.m.s.d. value of 0.279 Å. A comparison of Gan42B-WT with the other three homologous GH42 β-galactosidase structures reported to date indicates relatively high structural similarities, with r.m.s.d. values of 1.98, 1.21 and 1.87 Å for A4-β-gal (Hidaka *et al.*, 2002), Bca-β-gal (Maksimainen *et al.*, 2012) and *Bl*Gal42A (Viborg *et al.*, 2014), respectively (Fig. 2*c*). All four proteins are built of three relatively similar domains per monomer; however, although domain A seems to be closely similar between these structures, domains B and C seem to vary more significantly in their general conformation and in some local loop structures (Fig. 2*c*). Nonetheless, despite slight differences in the number and the exact orientation of their secondary-structure elements, the general topology of these four proteins is practically identical.

3.2. The Gan42B trimer

As mentioned, the crystallographic asymmetric unit of Gan42B contains three protein monomers (chains *A*, *B* and *C*) that form a compact homotrimer with near-C_3 noncrystallographic symmetry (Fig. 3). The overall shape of this homotrimer resembles a flowerpot, in which each subunit interacts with the other two to form a cone-shaped tunnel cavity in the centre of the flowerpot structure (Fig. 3*a*). The dimensions of the cone tunnel are ∼35 Å in diameter for the wide opening (top), ∼5 Å in diameter for the small opening (bottom) and ∼40 Å in length (top to bottom). The pseudo-C_3 symmetry axis passes vertically through the centre of the tunnel. The three active sites of Gan42B are located near the small opening of the cone tunnel (10–15 Å from the bottom), all facing the centre of the trimer (Fig. 3*a*). The active sites are

situated at the interfaces between the monomers, so that every two neighbouring monomers participate in the formation of each of the three complete active sites (see §3.3). Electrostatic surface calculations with the *APBS* plugin (Unni *et al.*, 2011) reveal that the entrances to the active sites possess a significant negative potential (Figs. 3*b* and 3*c*), as frequently

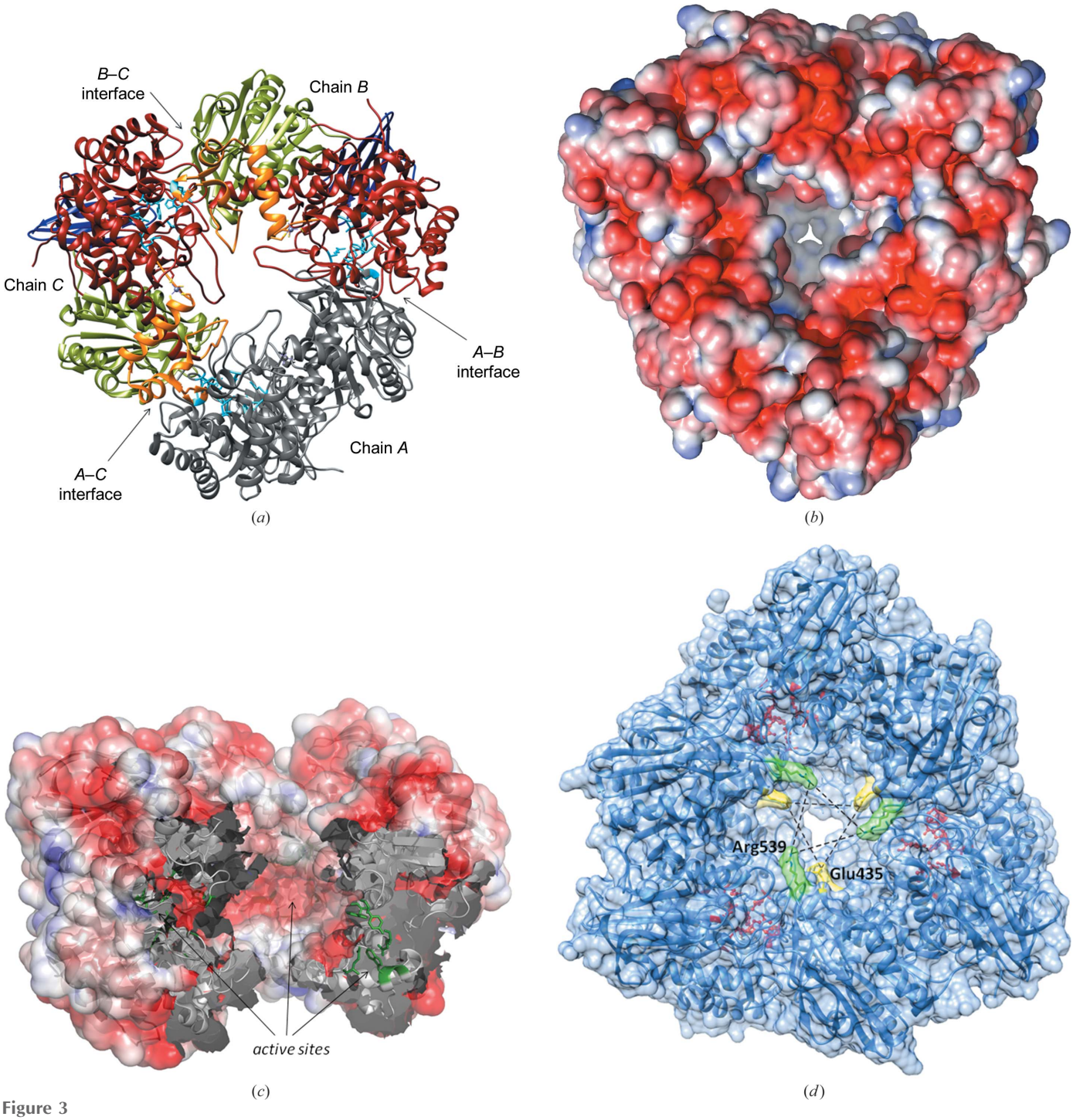


Figure 3
The three-dimensional structure of the Gan42B trimer. (*a*) Ribbon representation of the homotrimer. Chain *A* is coloured grey and chains *B* and *C* according to the different domains: domain A in red, domain B in green, domain C in dark blue and subdomain H in orange. The residues participating in the active sites are coloured light blue. (*b*) Electrostatic surface representation of the Gan42B trimer (top view) calculated with the *APBS* plugin (Unni *et al.*, 2011) as implemented in *PyMOL* (DeLano, 2002). Red represents negative potential and blue represents positive potential. The local potential value at each point is reflected by the colour shade, where the gradient scale is in the range from $4kT/e$ (dark blue) to $-4kT/e$ (dark red). (*c*) Similar electrostatic surface presentation of the Gan42B trimer (side view), where monomer *C* has been removed to allow a view into the internal cavity. This figure shows the 'funnel shape' of the central cavity and the relative concentration of negative potential around the active sites. (*d*) Surface representation of the Gan42B trimer (top view), demonstrating the involvement of Arg539 (green) and Glu435 (yellow) in forming the 'size-selectivity filter' in the bottom part of the cavity. The residues participating in the active sites are shown in red.

Table 3
The monomer–monomer interactions in the Gan42B trimer.

Only the interactions at the *A*–*B* interface are presented. The other interfaces contain similar symmetry-related interactions.

Monomer *A* (domains B/H)		Monomer *B* (domain A)			
Residue	Atom	Residue	Atom	Distance (Å)	Type of interaction
Trp193	NE1	Gln26	OE1	3.0	Hydrogen bond
	CZ2	Trp27	CZ2	3.8	π-Stacking
Trp197	CH2	Phe370	CD2	3.9	π-Stacking
Ser198	OG	Ser57	N	2.9	Hydrogen bond
His199	NE2	Gly119	O	2.7	Hydrogen bond
Ser205	OG	Gln100	OE1	2.7	Hydrogen bond
Gln206	NE2		OE1	3.5	Hydrogen bond
Ile213	O	Leu116	N	2.8	Hydrogen bond
Glu215	OE1	Gly119	N	3.4	Hydrogen bond
	OE2			3.1	Hydrogen bond
Thr217	O	Ser369	OG	3.4	Hydrogen bond
Phe432	O	Asn335	N	2.8	Hydrogen bond
			ND2	2.9	Hydrogen bond
	CE2	Phe373	CE2	4.2	π-Stacking
Tyr478	CE1			3.6	π-Stacking
Ala434	N	Lys333	O	3.3	Hydrogen bond
Glu435	OE2		NZ	3.2	Salt bridge
Asp511	OD2	Lys372	NZ	2.6	Salt bridge
Ala513	N	Arg367	O	3.0	Hydrogen bond
	O		NE	2.8	Hydrogen bond
Gly516	N	His380	O	2.7	Hydrogen bond
Glu529	OE1	Arg383	NH2	2.8	Salt bridge
Glu532	OE1	Arg387	NE	2.8	Salt bridge
Asp534	OD1	Arg339	NH2	2.9	Salt bridge
	OD2			2.7	Salt bridge
Asp574	OD2			3.3	Salt bridge
Tyr537	CE2	Trp294	CZ3	3.5	π-Stacking
Arg539	N	Asp293	OD2	2.9	Hydrogen bond
	NH1		O	3.1	Hydrogen bond
Asp540	OD2	Lys336	NZ	2.7	Hydrogen bond

observed in other glycoside hydrolases (Lansky, Salama, Solomon *et al.*, 2014).

The Gan42B trimer is held together by a number of hydrogen bonds, salt bridges and π-stacking interactions (Table 3). Each monomer interacts with its two neighbouring monomers *via* two main contact areas. On one side of the monomer this contact area involves the C-terminal loops of the TIM-barrel of domain A, and on the other side of the monomer this contact area involves parts of domain B and subdomain H (Fig. 3*a*). The monomer–monomer interactions forming the trimer can thus be categorized into three related contact interfaces, namely the *A*–*B* interface (the contact area between monomers *A* and *B*), the *B*–*C* interface and the *C*–*A* interface, each formed by the C-terminal loops of domain A of one monomer and domains B and H of the neighbouring monomer. Although slightly different in their detailed interatomic distances, these interfaces are quite similar and are generally related to each other by the pseudo-C_3 symmetry axis of the trimer. The specific interactions forming one such interface (the *A*–*B* interface) are summarized in Table 3, based on the final structure of Gan42B-WT. These intermolecular contacts involve a relatively large number of interactions, specfically 18 hydrogen bonds, seven salt bridges and five π-stacking interactions. The trimeric assembly observed in the current crystal structure of Gan42B therefore seems to be highly stable and not easily disrupted, making it more likely to be of high biological significance.

Gel-permeation chromatographic analysis of Gan42B, conducted in solutions closely similar to the corresponding physiological environment, demonstrates that the protein is present in solution practically in a single oligomeric form with a molecular weight of about 235 kDa (Fig. 1), conforming well with the homotrimer assembly observed in the current crystal structure. Theoretical calculations conducted with *PISA* (Krissinel & Henrick, 2007) predict a free energy of dissociation ($\Delta G^{\rm diss}$) value of 72.3 kcal mol^{-1} upon dissociation of the trimeric assembly, implying considerably strong assembly interactions. In addition, a similar trimeric assembly has also been observed in the crystal structures of the three other homologous GH42 β-galactosidases reported to date (Hidaka *et al.*, 2002; Maksimainen *et al.*, 2012; Viborg *et al.*, 2014). The significance of such observations is strengthened considering the different crystallographic environments of these four structures, with one trimer in the crystallographic asymmetric unit of Gan42B and *Bl*Gal42A (Viborg *et al.*, 2014), one monomer in the crystallographic asymmetric unit of A4-β-Gal (Hidaka *et al.*, 2002) and two trimers in the crystallographic asymmetric unit of Bca-β-Gal (Maksimainen *et al.*, 2012). Interestingly, although the residues involved in the interaction between the three monomers of Gan42B are not conserved in the homologous proteins, their interactions take place in similar regions of the enzymes, indicating that the intersubunit interactions are formed by the same structural elements. Thus, gel filtration, theoretical calculations and comparison with homologous GH42 enzymes all confirm that the trimeric quaternary assembly of Gan42B, observed here in the crystal structure, is also likely to be the biologically relevant form of the protein.

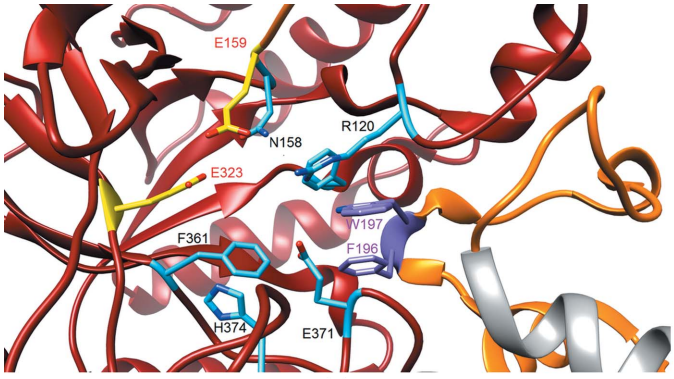
Although various oligomeric assemblies are increasingly being observed and reported in GH structures (Lansky, Alalouf, Solomon *et al.*, 2014; Lansky, Salama, Solomon *et al.*, 2014), the reasons for such oligomerizations are not always completely clear. In the particular case of Gan42B presented here, the specific oligomeric assembly observed is obviously driven primarily by the catalytic activity, as the full active sites are formed in the monomer–monomer interfaces created by such an assembly. However, the oligomerization here may also be the evolutionary result of the development of increased stability. This may be one of the contributing factors to the high thermal stability of the protein, which functions optimally at 65°C (Tabachnikov & Shoham, 2013). The trimeric oligomerization of Gan42B may also contribute to the increased selectivity and catalytic activity of the protein. Since the narrow opening of the central tunnel is only ∼5 Å in diameter, which is too narrow for the entrance of a sugar substrate, the substrate probably passes through the wider opening (∼35 Å) in order to reach the active sites of the protein. This tunnel, however, significantly narrows down during descent from the wider opening to the active site, and therefore seems to serve as a form of 'size-selectivity filter'. For example, Arg539 and its other two symmetry mates, which are situated in the tunnel lining, are at a distance of 15.8 Å from one another. A similar

case occurs with Glu435 and its symmetry mates (situated slightly below Arg539), which are at a distance of 16.4 Å from each other (Fig. 3*c*). These arrangements seem to enable only substrates of suitable size to traverse the tunnel and reach the active sites, an arrangement that is also suitable for the exo mode of activity observed in Gan42B (Tabachnikov & Shoham, 2013). Another possible reason for the trimeric assembly of Gan42B can be related to the 'molecular crowding' effect, where the substrate has more active-site pockets to choose from once it enters the central cavity of the protein, increasing the statistically successful encounters between the incoming substrate and the catalytic sites. Thus, the observed trimerization of Gan42B could be mainly rationalized by proper catalytic activity, but additionally by increased stability, size-based substrate selectivity, and protection from unwanted and unregulated hydrolysis.

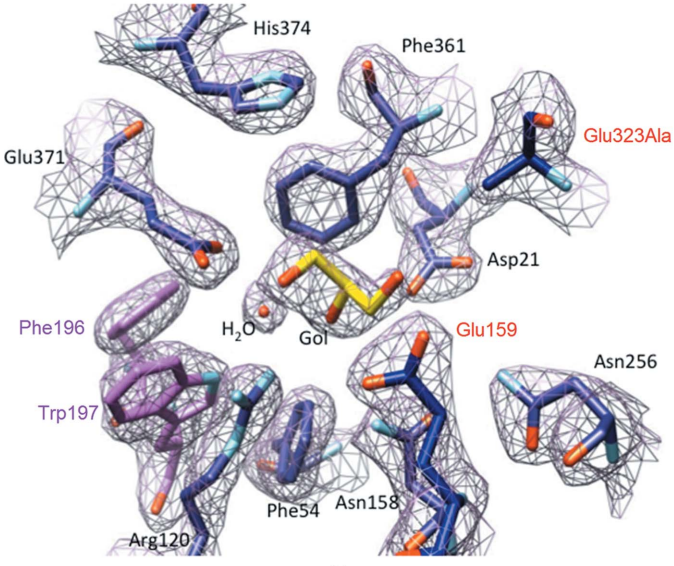
3.3. The active site

The catalytic active sites of Gan42B are located within the TIM-barrel fold (domain A) of each of the protein monomers and are complemented by a small part of subdomain H (residues Phe196 and Trp197). This structural complementation of the active site is provided by the adjacent monomer across the monomer–monomer interface within the Gan42B trimer (Fig. 4*a*). The active site is of a 'pocket' topology (Davies & Henrissat, 1995), suitable for the exo mode of action of Gan42B. The catalytic residues, Glu159 (acid/base) and Glu323 (nucleophile), are positioned at the two sides of the pocket at a distance of 5.3 Å from each other, confirming the prediction (based on sequence homology) that Gan42B acts *via* a retaining mechanism (Davies & Henrissat, 1995).

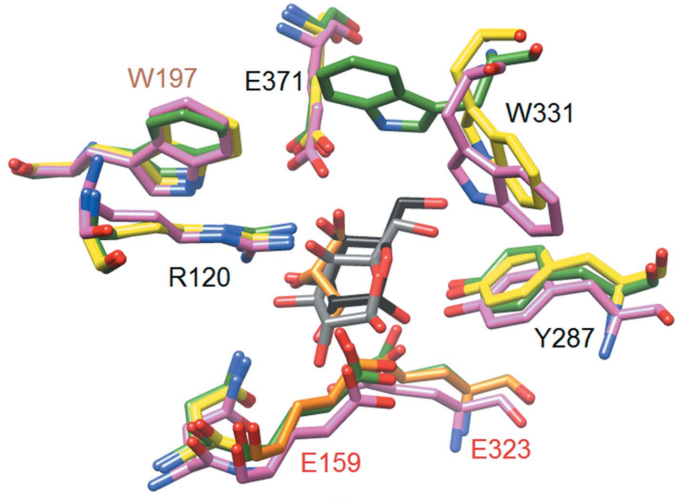
A glycerol molecule originating from the cryoprotecting solution used for crystallographic data collection was found in the active site of each of the Gan42B monomers of the trimer (Fig. 4*b*). The residues that interact directly with this bound glycerol are Glu371, Asn158, Arg120, Phe54 and Phe361, the catalytic residues Glu323 and Glu159, and Asp21, which forms a water-mediated interaction with the O2 atom of the active-site glycerol. Since glycerol often mimics partial structures of sugars owing to similarly positioned chemical functional groups (Solomon *et al.*, 2007; Lansky, Salama, Solomon *et al.*, 2014), these residues are probably also involved in interaction with real substrates of Gan42B, especially considering their



(*a*)



(*b*)



(*c*)

Figure 4
The active site of Gan42B. (*a*) An enlargement of the H–A domain interface, demonstrating how subdomain H (orange, right) complements the active site situated mainly in domain A (red, left). Residues belonging to subdomain H are shown in purple, residues belonging to domain A are shown in blue, catalytic residues are shown in yellow and the grey ribbon represents domain A of the adjacent monomer. (*b*) The active site of Gan42B-E323A together with a bound glycerol molecule (Gol; yellow), as modelled in the final electron-density map ($2F_o - F_c$) at 2.5 Å resolution, contour level of $\sim 1\sigma$. The acid/base (Glu159) and the mutated nucleophile (E323A) are labelled in red, and Trp197 and Phe196, which originate from the symmetry-related monomer, are shown in purple. (*c*) Superposition of the active sites of Gan42B (yellow) with the active sites of the GH42 homologues Bca-β-gal (green; PDB entry 3tty; Maksimainen *et al.*, 2012) and *Bl*Gal42A (pink; PDB entry 4uni; Viborg *et al.*, 2014). The glycerol molecule found in Gan42B is shown in orange, superimposed on the corresponding bound products in the homologous structures (α-galactose, PDB entry 3tty, black; β-galactose, PDB entry 4uni, grey). The catalytic residues of Gan42B are labelled in red (and shown in orange) and Trp197 from the symmetry-related adjacent monomer is labelled in orange (top left).

high sequence conservation (Fig. 5). Indeed, a comparison between the active site of Gan42B and those of the three other GH42 β-galactosidases of known structure (in complex with a galactose product; Hidaka *et al.*, 2002; Maksimainen *et al.*, 2012; Viborg *et al.*, 2014) further confirms the critical involvement of these residues in Gan42B substrate (and product) binding (Fig. 4*c*). Such comparison also indicates that the conserved aromatic residues His374, Trp331, Tyr287 and Trp197 are likely to be involved in substrate binding through π-stacking interactions. Fig. 4(*c*) shows that the glycerol molecule trapped in the active site of Gan42B indeed quite closely mimics half of a bound galactose molecule in the active site, demonstrating the high binding affinity of β-galactosidases such as Gan42B and its homologues towards properly positioned hydroxyl groups of sugars and sugar-like molecules. All five hydroxyl groups of the bound galactose molecule (in the corresponding cases of the Gan42B homologues) are hydrogen-bonded to at least one active-site protein residue, indicating that the binding of a galactosyl moiety to the active site of these enzymes is relatively strong and that the relative position of each of the sugar

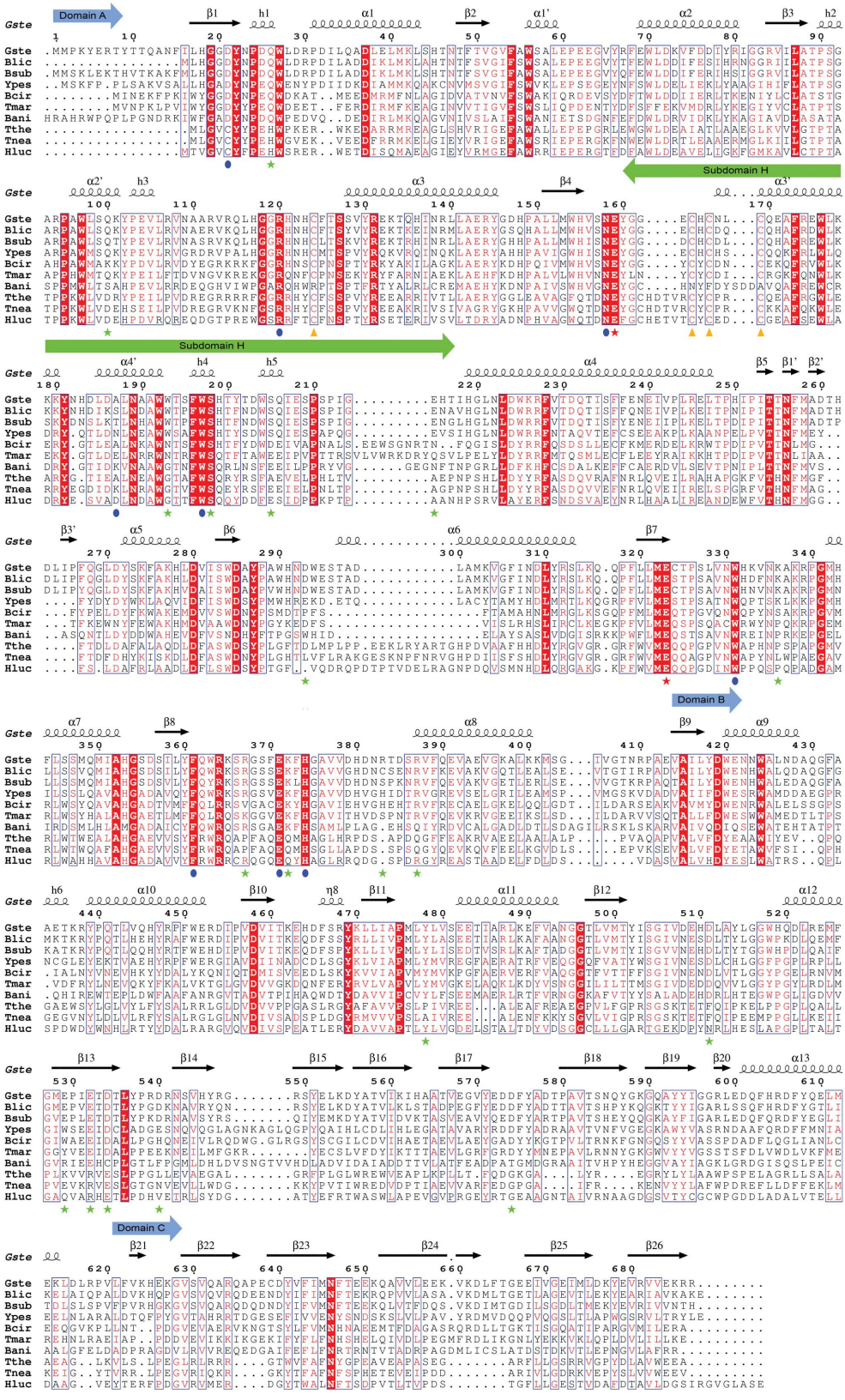


Figure 5
Homology sequence alignment of Gan42B. Amino-acid sequence alignment of Gan42B and related GH42 β-galactosidases. The homologous proteins are from the following sources: Gste, *Geobacillus stearothermophilus* (Gan42B); Blic, *Bacillus licheniformis*; Bsub, *Bacillus subtilis*; Ypes, *Yersinia pestis*; Bcir, *Bacillus circulans* sp. *alkalophilus*; Tmar, *Thermotoga maritima*; Bani, *Bifidobacterium animalis* subsp. *lactis* Bl-04; Tthe, *Thermus thermophilus* A4; Tnea, *Thermotoga neapolitana*; Hluc, *Haloferax lucentense*. Strictly conserved residues are boxed in red. Highly similar sequence regions are marked in red letters and boxed in blue. The beginnings of domains A, B and C are indicated with light-blue arrowheads. Subdomain H is indicated with light-green arrowheads (within domain A). The catalytic residues are marked with red stars. Residues involved in galactose binding, intersubunit interactions and metal binding are denoted by blue dots, green stars and orange triangles, respectively. The sequence numbering and secondary-structure elements correspond to the sequence of Gan42B.

hydroxyl groups is important. Moreover, considering the high structural similarities of the actives sites of the reported GH42 β-galactosidases, it was possible (and meaningful) to model a bound galactose product in the active site of Gan42B-WT (data not shown). This model confirms that similarly to the case of the three homologues, a bound galactose also fits well in the active site of Gan42B and all of its hydroxyls are involved in enzyme–product interactions. In addition to these multiple hydroxyl hydrogen bonds, several aromatic π-stacking interactions with the bound galactose are formed by Phe54, Tyr287, Trp331, Phe361 and Trp197 (Fig. 4*c*), enzyme–ligand interactions of the types that are commonly observed amongst glycoside hydrolases (Solomon *et al.*, 2007; Lansky, Salama, Solomon *et al.*, 2014).

As mentioned above, the highly conserved Trp197 belongs to the adjacent symmetry-related monomer and seems to complement the active-site pocket of Gan42B (Fig. 4*a*). To examine the role of this residue in enzyme binding and catalysis, mutants of Trp197 (Gan42B-W197H and Gan42B-W197G) have been prepared and examined. The mutations functional tests showed that the catalytic activity of Gan42B-W197H at 8 m*M* *p*NPG was 41% of the activity of the wild-type enzyme, whereas the catalytic activity of Gan42B-W197G was only 12% of the wild-type activity. These results confirmed that Trp197 plays an important role in the catalytic mechanism of Gan42B, probably *via* substrate binding and/or recognition.

Interestingly, although the active sites of the four reported GH42 β-galactosidases appear to be considerably similar (Fig. 4*c*), a significant conformational difference is apparent in the position and orientation of the Trp331 side chain. The indole group of this residue moves significantly between the different structures, pointing to one side of the active-site pocket in the current structure of Gan42B (Fig. 4*c*, yellow) and the structure of *Bl*Gal42A (pink), while pointing to the opposite side of the pocket in the structure of Bca-β-gal (green). This residue is situated at the entrance to the active-site pocket and as such may be involved in aromatic π-stacking interactions with an incoming substrate. The location and conformational flexibility of this residue suggests that Trp331 is probably involved in transient pre-catalytic binding of the substrate, helping to guide it into the active site of the enzyme (Maksimainen *et al.*, 2012).

Another point of potential mechanistic interest may be related to the configuration of the trapped galactose molecules in the active sites of these homologous β-galactosidases. The bound galactose in the reported structures of A4-β-gal (Hidaka *et al.*, 2002) and Bca-β-gal (Maksimainen *et al.*, 2012) adopts the α-configuration, whereas the bound galactose in *Bl*Gal42A (Viborg *et al.*, 2014) adopts the β-configuration, differing in the specific configuration of the O1 atom of the sugar ring. The finding of α-galactose in the active sites of A4-β-gal and Bca-β-gal is quite unexpected, since these enzymes hydrolyze only galacto-oligosaccharides with a β-configuration and not with an α-configuration, at least in the case of substrates with two or more sugar units. GH42 enzymes, however, operate through a retaining mechanism, which involves an attack by a nucleophile, resulting in an intermediate state that is covalently bound to the nucleophile. Such an intermediate, with an inverted configuration with respect to the original substrate, is then converted back to the original configuration following another nucleophilic attack by a nearby water molecule (Davies & Henrissat, 1995; Lansky, Salama, Solomon *et al.*, 2014). The α-galactose molecule found in the active site of the reported structures, of opposite configuration to the original β-configuration, may thus mimic the intermediate state. If this was indeed the case, however, we would have expected to find the nucleophile closer to the α-galactose rather than to the β-galactose configuration, as the intermediate state involves the nucleophile covalently bound to the sugar in an α-configuration. Instead, as seen in Fig. 4(*c*), the bound β-galactose is closer to the nucleophile and is shifted $\sim$0.7 Å with respect to the α-galactose molecule. This discrepancy, however, should be considered in light of the experimental resolution limits (2.2–2.6 Å), which do not always permit conclusive assignments of exact locations, configurations and conformations.

3.4. The zinc-binding site

Four cysteine residues, Cys124, Cys164, Cys166 and Cys169, form a metal-binding cluster that is located in domain A of Gan42B (Figs. 2*a*, 6*a* and 6*b*). These residues are relatively conserved within the GH42 family (Fig. 5). Cys164, Cys166 and Cys169 are positioned at the root of subdomain H, whereas Cys124 is positioned in the TIM-barrel domain (Fig. 2*a*). Inductive coupled plasma (ICP) measurements revealed the presence of 1.12 Zn atoms per Gan42B monomer, and fractional amounts of Mg and Ca atoms were also detected (data not shown). Al, Cd, Co, Cu, Fe, Mn, Ni and Pb atoms were not detected. The metal ion in the observed metal cluster was therefore modelled as a zinc cation, as confirmed by its crystallographic refinement, which resulted in a very reasonable average *B* factor of 13.8 Å^2. The slight excess of Zn ions and residual Mg and Ca ions detected by the ICP analysis are likely to be the result of nonspecific binding, some of which can potentially take place in the His-tag peptide at the N-terminus of the protein, which was not removed for all experiments. Biochemical experiments show that the presence of various metal ions (1 m*M* Co, Fe, Mg, Mn or Ca) does not significantly affect the activity of Gan42B towards *p*NPG, confirming that these metal ions are not necessary for the enzymatic activity of Gan42B. Addition of the chelating agent EDTA (10 m*M*) did not affect the catalytic activity as well. ICP analysis, however, shows that Gan42B contains 0.95 Zn atoms per monomer after the addition and removal of 10 m*M* EDTA. These results indicate that the coordination complex of the metal with the four Cys ligands is quite strong, making removal of the zinc ion by EDTA quite difficult. The addition of 1 m*M* Zn to the EDTA-treated Gan42B did not change the catalytic parameters, confirming that only one zinc ion per protein monomer is functionally required.

Mutations at the zinc-binding site were performed in order to examine whether the zinc ion of the zinc-binding site affects the catalytic activity of Gan42B. Two mutants were prepared

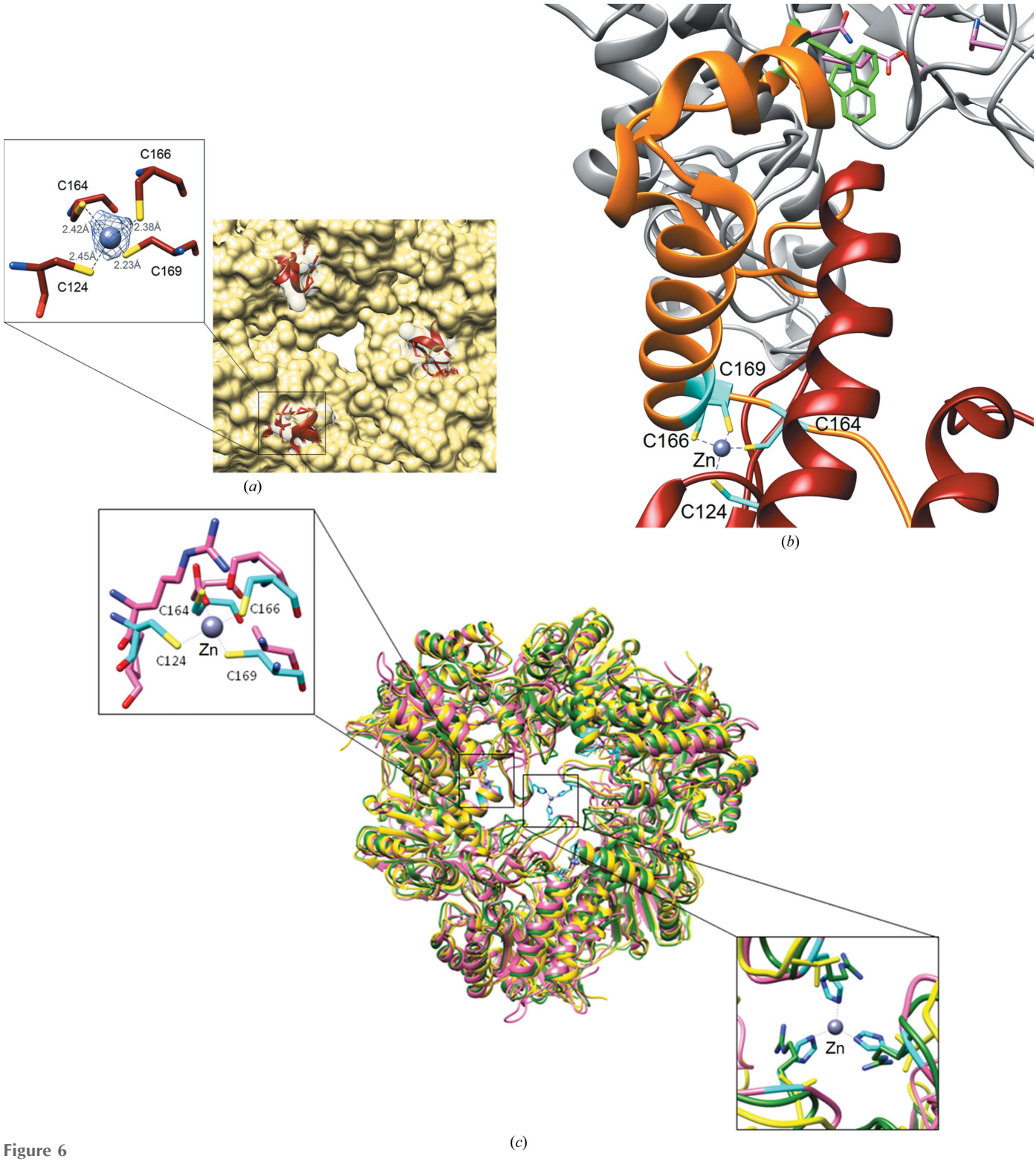


Figure 6
The zinc-binding site of Gan42B. (*a*) The locations of the three zinc-binding sites (marked in red) in the Gan42B trimer (viewed into the narrower 'bottom' of the central cavity). The inset shows an enlargement of the zinc-binding site (with the corresponding difference electron density around the Zn atom) and the tetrahedral coordination formed by the four Cys residues (Zn–S distances in Å). (*b*) Enlargement of the interface between subdomain H (orange) and domain A (red), demonstrating the key role of the zinc cluster (turquoise, bottom) in stabilizing the conformation of subdomain H, which allows it to form the complete active site together with the adjacent Gan42B monomer (grey). The active-site residues belonging to subdomain H are shown in green; the active-site residues belonging to the adjacent monomer are shown in pink (top). (*c*) Superposition of the Gan42B trimer (yellow) on the trimers of two other GH42 homologues, Bca-β-gal (green; PDB entry 3tty; Maksimainen *et al.*, 2012) and *Bl*Gal42A (pink; PDB entry 4uni; Viborg *et al.*, 2014), demonstrating the differences in location and coordination of the zinc clusters. In Gan42B, A4-β-Gal and Bca-β-gal the bound zinc is coordinated by four Cys residues (turquoise), while in *Bl*Gal42A these residues are missing (upper left inset). In *Bl*Gal42A the zinc is coordinated by three His residues (turquoise) that are missing from the other structures (bottom right inset).

in which the Cys residues were replaced by Ala: a single mutant C124A and a double mutant C164A/C166A. Surprisingly, the catalytic activity of the single mutant (at 8 m*M* *p*NPG, 40°C) was 194% relative to the wild-type enzyme, whereas the double mutant exhibited only 36% of the wild-type activity. The contribution of the metal cluster to protein thermal stability in the wild-type and mutant proteins was tested by differential scanning calorimetry (DSC). These DSC experiments indicated that the melting temperature of the C164A/C166A double mutant was almost 10°C lower than of the wild-type protein, at 67.5 and 77.0°C for the mutant and wild-type proteins, respectively (Fig. 7).

Since the zinc cluster is relatively distant from the active site (about 20 Å), these effects on catalysis and thermal stability can potentially result from either indirect structural changes of the active and/or binding sites, or general effects on the overall stability of the protein. A closer look at the current Gan42B structure reveals that the metal-binding cluster is located at the root of subdomain H, at its branching position from domain A, holding together the main helix of this subdomain with a key loop of domain A (Figs. 2*a* and 6*b*). Hence, the zinc-coordination complex may be involved in rigidifying subdomain H, thereby influencing its proper orientation and the exact positioning of the Trp197 residue situated in the tip of this domain. The side chain of Trp197 forms a critical part of the active site of Gan42B (as mentioned in §3.3 above), and the presence of the zinc cluster may therefore stabilize it in a catalytically suitable conformation. Thus, the different activity parameters observed for mutants of the zinc-coordinating residues may reflect a slight shift of the exact position of Trp197, which in turn can affect the catalytic activity of Gan42B. These changes in catalysis may be manifested by even relatively small changes in the active-site envelope, thereby altering the exact binding orientation of the substrate, intermediates and products. Such changes may therefore lead to different enzyme–ligand interactions along the reaction coordinate, resulting in the reduced activity observed for the C164A/C166A double mutant and increased activity observed for the C124A single mutant. While usually unexpected for the natural substrate at physiological conditions, the significantly increased catalytic activity of the Gan42B/C142A mutant could be a result of the particular synthetic substrate tested and the non-natural reaction conditions used, and as such do not reflect a physiologically relevant effect on the catalytic activity.

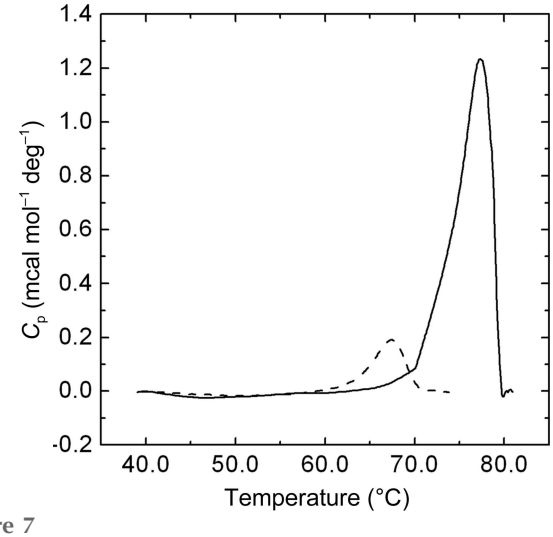


Figure 7
Thermal stability of Gan42B. Heat-capacity curves for Gan42B-WT (solid line) and the C164A/C166A mutant (dashed line). These curves show that the wild-type protein has a relatively high melting temperature (T_m = 77.0°C), but this value is decreased significantly by mutation of two of the zinc-coordinating cysteine ligands (T_m = 67.5°C), demonstrating the key contribution of the zinc–protein interactions to the overall stabilization of the Gan42B protein.

An alternative explanation for the catalytic behaviour of the mutants of the zinc-binding cysteines could be related to the fact that two of these cysteine residues (Cys164 and Cys166) are located on the same polypeptide segment as Glu159, the catalytic acid/base (Fig. 2*a*). In this respect, even small changes in the orientation and rigidity of this key segment can easily influence the exact position and/or orientation of Glu159, thereby modifying the corresponding catalytic parameters. These indirect conformational changes can therefore account for either an increase or a decrease in the overall catalytic profile of Gan42B, as observed in the cases of the C124A and C164A/C166A mutants, respectively. Nevertheless, to obtain more meaningful conclusions concerning the role of the zinc, and the influence of the zinc ligands on the catalytic activity of Gan42B, these aspects should be further investigated using a systematic series of zinc-coordination mutants tested with natural substrates under physiologically relevant reaction conditions.

The presence of a 'structural' zinc ion in Gan42B is in agreement with similar assignments in the structures of the homologous GH42 *β*-galactosidases A4-*β*-gal and Bca-*β*-gal (Hidaka *et al.*, 2002; Maksimainen *et al.*, 2012). Zinc ions were modelled in the corresponding positions of these enzymes, also coordinated by four Cys residues. Surprisingly, however, in the structure of the homologous GH42 *Bl*Gal42A (32% identity to Gan42B; Viborg *et al.*, 2014) a zinc-binding site was not found in the same place, but in a totally different position. In this enzyme, the zinc ion is located in the centre of the trimer, connecting the three monomers together at the centre of the narrow opening of the internal tunnel, and coordinated to three His residues rather than to four Cys residues (Fig. 6*c*). As can be seen in Fig. 5, these His residues are not conserved in Gan42B and the two homologous GH42 *β*-galactosidases mentioned above, while conversely, *Bl*Gal42A does not contain the four conserved Cys residues present in the other three proteins. In the cases of Gan42B, A4-*β*-gal and Bca-*β*-gal the zinc ion appears to indirectly strengthen the trimeric assembly (and the active site formed by this trimerization) by stabilizing the proper conformation of subdomain H and its proper intermolecular interactions with the neighbouring monomers (see above). In the case of *Bl*Gal42A, however, the zinc ion plays a more direct role in such trimer stabilization, this time by its coordination to three histidine residues, each from a different monomer of the trimer. Despite these differences, the active sites of all four homologues are formed

similarly at the interface between neighbouring monomers of the trimer. Thus, in all four cases the zinc ion appears to stabilize the active sites of the enzymes by stabilization of the trimeric formation, either directly or indirectly.

In summary, the homotrimer assembly observed here in the crystal structure of Gna42B, a common feature of the GH42 β-galactosidases reported to date, seems to be critical for both the function and stability of the enzyme. This key quaternary structure is stabilized, among other things, by zinc ions and their interactions with specific protein residues; however, the exact role of these metal ions in catalysis, if any, is not yet completely clear. These functional aspects should be further investigated by dedicated and systematic mutagenesis studies, some of which are currently under way in our laboratory.

Acknowledgements

This work was supported by Israel Science Foundation Grants 500/10, 152/11 and 1905/15, the I-CORE Program of the Planning and Budgeting Committee, The Israeli Ministry of Environmental Protection, The Israeli Ministry of Science and the Grand Technion Energy Program (GTEP). This work 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. This work was also supported by internal funds of the Hebrew University of Jerusalem and a Klein PhD fellowship to HVS. SL is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. We thank the staff at the European Synchrotron Research Facility (ID14-1 and BM14 beamlines) and the Diamond Light Source Facility (I04 beamline) and the research group of N. C. Chayen (Imperial College, England) for their helpful support in X-ray synchrotron data measurement and analysis. YS holds the Erwin and Rosl Pollak Chair in Biotechnology at the Technion.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Alalouf, O., Balazs, Y., Volkinshtein, M., Grimpel, Y., Shoham, G. & Shoham, Y. (2011). *J. Biol. Chem.* **286**, 41993–42001.
Almog, O., Greenblatt, H. M., Spungin, A., Ben-Meir, D., Blumberg, S. & Shoham, G. (1993). *J. Mol. Biol.* **230**, 342–344.
Almog, O., Klein, D., Braun, S. & Shoham, G. (1994). *J. Mol. Biol.* **235**, 760–762.
Bar, M., Golan, G., Nechama, M., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2004). *Acta Cryst.* D**60**, 545–549.
Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J. L. S. & Subramaniam, S. (2015). *Science*, **348**, 1147–1151.
Ben-David, A., Bravman, T., Balazs, Y. S., Czjzek, M., Schomburg, D., Shoham, G. & Shoham, Y. (2007). *Chembiochem*, **8**, 2145–2151.
Ben-David, A., Shoham, G. & Shoham, Y. (2008). *Chem. Biol.* **15**, 546–551.
Bravman, T., Zolotnitsky, G., Belakhov, V., Shoham, G., Henrissat, B., Baasov, T. & Shoham, Y. (2003). *Biochemistry*, **42**, 10528–10536.
Bravman, T., Zolotnitsky, G., Shulami, S., Belakhov, V., Solomon, D., Baasov, T., Shoham, G. & Shoham, Y. (2001). *FEBS Lett.* **495**, 39–43.
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.
Champion, E., Remaud-Simeon, M., Skov, L. K., Kastrup, J. S., Gajhede, M. & Mirza, O. (2009). *Acta Cryst.* D**65**, 1309–1314.
Cowtan, K., Emsley, P. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 233–234.
Daniel, R. A., Haiech, J., Denizot, F. & Errington, J. (1997). *J. Bacteriol.* **179**, 5636–5638.
Dann, R., Lansky, S., Lavid, N., Zehavi, A., Belakhov, V., Baasov, T., Dvir, H., Manjasetty, B., Belrhali, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* F**70**, 1675–1682.
Davies, G. & Henrissat, B. (1995). *Structure*, **3**, 853–859.
Delangle, A., Prouvost, A.-F., Cogez, V., Bohin, J.-P., Lacroix, J.-M. & Cotte-Pattat, N. H. (2007). *J. Bacteriol.* **189**, 7053–7061.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Durand, P., Lehn, P., Callebaut, I., Fabrega, S., Henrissat, B. & Mornon, J.-P. (1997). *Glycobiology*, **7**, 277–284.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Gat, O., Lapidot, A., Alchanati, I., Regueros, C. & Shoham, Y. (1994). *Appl. Environ. Microbiol.* **60**, 1889–1896.
Gilboa, R., Bauer, A. J. & Shoham, G. (1998). *Acta Cryst.* D**54**, 1467–1470.
Golan, G., Shallom, D., Teplitsky, A., Zaide, G., Shulami, S., Baasov, T., Stojanoff, V., Thompson, A., Shoham, Y. & Shoham, G. (2004). *J. Biol. Chem.* **279**, 3014–3024.
Golan, G., Zharkov, D. O., Fernandes, A. S., Zaika, E., Kycia, J. H., Wawrzak, Z., Grollman, A. P. & Shoham, G. (2004). *Acta Cryst.* D**60**, 1476–1480.
Henrissat, B. & Davies, G. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.
Hidaka, M., Fushinobu, S., Ohtsu, N., Motoshima, H., Matsuzawa, H., Shoun, H. & Wakagi, T. (2002). *J. Mol. Biol.* **322**, 79–91.
Hinz, S. W. A., van den Broek, L. A. M., Beldman, G., Vincken, J. P. & Voragen, A. G. J. (2004). *Appl. Microbiol. Biotechnol.* **66**, 276–284.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Hondoh, H., Kuriki, T. & Matsuura, Y. (2003). *J. Mol. Biol.* **326**, 177–188.
Hövel, K., Shallom, D., Niefind, K., Baasov, T., Shoham, G., Shoham, Y. & Schomburg, D. (2003). *Acta Cryst.* D**59**, 913–915.
Hövel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. & Schomburg, D. (2003). *EMBO J.* **22**, 4922–4932.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Lansky, S., Alalouf, O., Salama, R., Dvir, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* F**70**, 476–481.
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.
Lansky, S., Alalouf, O., Solomon, H. V., Alhassid, A., Govada, L., Chayen, N. E., Belrhali, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* D**70**, 261–278.
Lansky, S., Salama, R., Dann, R., Shner, I., Manjasetty, B. A., Belrhali, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* F**70**, 1038–1045.
Lansky, S., Salama, R., Solomon, V. H., Belrhali, H., Shoham, Y. & Shoham, G. (2013). *Acta Cryst.* F**69**, 695–699.
Lansky, S., Salama, R., Solomon, H. V., Feinberg, H., Belrhali, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* D**70**, 2994–3012.
Lansky, S., Zehavi, A., Dann, R., Dvir, H., Belrhali, H., Shoham, Y. & Shoham, G. (2014). *Acta Cryst.* F**70**, 225–231.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Maksimainen, M., Paavilainen, S., Hakulinen, N. & Rouvinen, J. (2012). *FEBS J.* **279**, 1788–1798.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
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.
O'Connell Motherway, M., Fitzgerald, G. F. & van Sinderen, D. (2011). *Microb. Biotechnol.* **4**, 403–416.
Ohtsu, N., Motoshima, H., Goto, K., Tsukasaki, F. & Matsuzawa, H. (2014). *Biosci. Biotechnol. Biochem.* **62**, 1539–1545.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
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.
Powell, H. R., Johnson, O. & Leslie, A. G. W. (2013). *Acta Cryst.* D**69**, 1195–1203.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Rees, D. C., Johnson, E. & Lewinson, O. (2009). *Nature Rev. Mol. Cell Biol.* **10**, 218–227.
Reiland, V., Fundoiano-Hershcovitz, Y., Golan, G., Gilboa, R., Shoham, Y. & Shoham, G. (2004). *Acta Cryst.* D**60**, 2371–2376.
Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
Rose, P. W. *et al.* (2013). *Nucleic Acids Res.* **41**, D475–D482.
Schwab, C., Sørensen, K. I. & Gänzle, M. G. (2010). *Syst. Appl. Microbiol.* **33**, 300–307.
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.
Shipkowski, S. & Brenchley, J. E. (2006). *Appl. Environ. Microbiol.* **72**, 7730–7738.
Shulami, S., Gat, O., Sonenshein, A. L. & Shoham, Y. (1999). *J. Bacteriol.* **181**, 3695–3704.
Shulami, S., Raz-Pasteur, A., Tabachnikov, O., Gilead-Gropper, S., Shner, I. & Shoham, Y. (2011). *J. Bacteriol.* **193**, 2838–2850.
Shulami, S., Zaide, G., Zolotnitsky, G., Langut, Y., Feld, G., Sonenshein, A. L. & Shoham, Y. (2007). *Appl. Environ. Microbiol.* **73**, 874–884.
Solomon, H. V., Tabachnikov, O., Feinberg, H., Govada, L., Chayen, N. E., Shoham, Y. & Shoham, G. (2013). *Acta Cryst.* F**69**, 1114–1119.
Solomon, V., Teplitsky, A., Shulami, S., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2007). *Acta Cryst.* D**63**, 845–859.
Tabachnikov, O. (2012). PhD thesis. Technion, Haifa, Israel.
Tabachnikov, O. & Shoham, Y. (2013). *FEBS J.* **280**, 950–964.
Teplitsky, A., Feinberg, H., Gilboa, R., Lapidot, A., Mechaly, A., Stojanoff, V., Capel, M., Shoham, Y. & Shoham, G. (1997). *Acta Cryst.* D**53**, 608–611.
Teplitsky, A., Mechaly, A., Stojanoff, V., Sainz, G., Golan, G., Feinberg, H., Gilboa, R., Reiland, V., Zolotnitsky, G., Shallom, D., Thompson, A., Shoham, Y. & Shoham, G. (2004). *Acta Cryst.* D**60**, 836–848.
Teplitsky, A., Shulami, S., Moryles, S., Shoham, Y. & Shoham, G. (2000). *Acta Cryst.* D**56**, 181–184.
Teplitsky, A., Shulami, S., Moryles, S., Zaide, G., Shoham, Y. & Shoham, G. (1999). *Acta Cryst.* D**55**, 869–872.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Unni, S., Huang, Y., Hanson, R. M., Tobias, M., Krishnan, S., Li, W. W., Nielsen, J. E. & Baker, N. A. (2011). *J. Comput. Chem.* **32**, 1488–1491.
Viborg, A. H., Fredslund, F., Katayama, T., Nielsen, S. K., Svensson, B., Kitaoka, M., Lo Leggio, L. & Abou Hachem, M. (2014). *Mol. Microbiol.* **94**, 1024–1040.
Yang, T.-C., Hu, R.-M., Hsiao, Y.-M., Weng, S.-F. & Tseng, Y.-H. (2003). *J. Mol. Microbiol. Biotechnol.* **6**, 145–154.
Zaide, G., Shallom, D., Shulami, S., Zolotnitsky, G., Golan, G., Baasov, T., Shoham, G. & Shoham, Y. (2001). *Eur. J. Biochem.* **268**, 3006–3016.
Zein, F., Zhang, Y., Kang, Y., Burns, K., Begley, T. P. & Ealick, S. E. (2006). *Biochemistry*, **45**, 14609–14620.