Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure–specificity relationships of an intracellular xylanase from *Geobacillus stearothermophilus*

V. Solomon,[a] A. Teplitsky,[a] S. Shulami,[b] G. Zolotnitsky,[b] Y. Shoham[b]* and G. Shoham[a]*

[a]Department of Inorganic 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 and Institute of Catalysis Science and Technology, Technion–Israel Institute of Technology, Haifa 32000, Israel

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

Received 21 February 2007
Accepted 21 May 2007

PDB References: IXT6, 2q8x, r2q8xsf; IXT6 at 1.45 Å, 1n82.

Geobacillus stearothermophilus T-6 is a thermophilic Gram-positive bacterium that produces two selective family 10 xylanases which both take part in the complete degradation and utilization of the xylan polymer. The two xylanases exhibit significantly different substrate specificities. While the extracellular xylanase (XT6; MW 43.8 kDa) hydrolyzes the long and branched native xylan polymer, the intracellular xylanase (IXT6; MW 38.6 kDa) preferentially hydrolyzes only short xylo-oligosaccharides. In this study, the detailed three-dimensional structure of IXT6 is reported, as determined by X-ray crystallography. It was initially solved by molecular replacement and then refined at 1.45 Å resolution to a final R factor of 15.0% and an $R_{\rm free}$ of 19.0%. As expected, the structure forms the classical $(\alpha/\beta)_8$ fold, in which the two catalytic residues (Glu134 and Glu241) are located on the inner surface of the central cavity. The structure of IXT6 was compared with the highly homologous extracellular xylanase XT6, revealing a number of structural differences between the active sites of the two enzymes. In particular, structural differences derived from the unique subdomain in the carboxy-terminal region of XT6, which is completely absent in IXT6. These structural modifications may account for the significant differences in the substrate specificities of these otherwise very similar enzymes.

1. Introduction

Endo-β-1,4-xylanases (EC 3.2.1.8) hydrolyze the β-1,4-backbone of xylan, the second most abundant polysaccharide in plant cell walls. Xylan is a heterogeneous polymer consisting of a β-1,4-linked xylopyranose backbone substituted with various groups such as acetyl, L-arabinofuranose, D-glucuronic acid or 4-*O*-methylglucuronic acid. Xylans display a large diversity with respect to the degree of polymerization (DP) and the nature of branching and substitution. Owing to the complex chemical nature of plant xylan, its complete hydrolysis requires the synergistic action of several hydrolytic enzymes with diverse specificities and modes of action, including β-xylosidase, α-L-arabinofuranosidase, α-D-glucuronidase and acetyl xylan esterase (Shallom & Shoham, 2003; Beg *et al.*, 2001). Of these xylan-processing enzymes, the xylanases are the most extensively studied and characterized and the three-dimensional structures of a number of them have recently been reported. Based on amino-acid sequence similarity, xylanases have been grouped into two main glycoside hydrolase families, GH10 and GH11, which are continuously updated in the CAZY database (http://www.cazy.org/; Henrissat & Davies, 1997; Coutinho & Henrissat, 1999). In

addition, enzymes with xylanase activity are also found in families GH5, GH7, GH8 and GH43 (Collins *et al.*, 2005; Parkkinen *et al.*, 2004; Shallom & Shoham, 2003; Brennan *et al.*, 2004). Xylanases have a wide range of commercial applications. In the pulp and paper industry, they are used for bio-bleaching of Kraft pulp in order to minimize the use of chlorine-based compounds. In the food industry, they are used for processing flour and improving its texture. Xylanases are also used for the degradation of plant biomass to soluble sugars that can be converted to ethanol for potential use as a biofuel (Kulkarni *et al.*, 1999; Zaide *et al.*, 2001; Shallom & Shoham, 2003; Polizeli *et al.*, 2005).

A significant part of the knowledge available today on the structure–function relationships of xylanases results from crystallographic three-dimensional structures. About 15 different such structures of family 10 xylanases are presently available in the Protein Data Bank (PDB), some of which were determined in the form of their complexes with a substrate, a substrate analogue or an inhibitor. These reported crystal structures include enzymes from *Cellulomonas fimi* (in complex with aza-sugar inhibitors and fluorocellobioside; PDB codes 1exp and 1fh9), from *Cellvibrio japonicus* (formally *Pseudomonas cellulose*; in complex with xylopentaose; PDB code 1e5n), from *Cellvibrio mixtus* (in complex with xylopyranose and hexaose; PDB code 1uqy), from *Clostridium thermocellum* NCIB 10682; PDB code 1xyz), from the alkalophilic *Bacillus* sp. NG-27 (in complex with a short xylooligosaccharide; PDB code 2fgl), from *Emericella nidulans* (in complex with a protein inhibitor; PDB code 1ta3), from *Geobacillus stearothermophilus* T-6 (in complex with xylopentaose; PDB code 1r87) and from *Penicillium simplicissimum* BT2246 (including complexes with a decorated xylotriose and a series of complexes with small oligoxyloses with n = 2–5; PDB code 1b3z). Also available in the PDB are structures of xylanases from *Streptomyces halstedii* JM8 (PDB code 1nq6, no substrate), *S. lividans* (complexes with the 2-fluoro-xylobiosyl intermediate 2-fluorocellobioside; PDB codes 1e0v and 1e0x), *S. olivaceoviridis* E-86 (complexes with several xylo-oligosaccharides and arabinofuranosyl decorated xylotriose; PDB codes 1it0, 1isx and 1v6v), *Thermoascus aurantiacus* (xylo-oligosaccharide complexes; PDB code 1gor) and *Thermotoga maritima* (xylobiose complex; PDB code 1vbr).

The overall structure of these GH10 enzymes corresponds to an eightfold α/β-barrel (TIM-barrel) with a typical deep groove in the centre, allowing an '*endo*' type of action on the large polysaccharide backbone. All family 10 xylanases hydrolyze the glycosidic bond in a double-displacement 'retaining' mechanism using two catalytic acidic residues, where one residue acts as a nucleophile (base) and the other acts as a general acid/base. The binding of oligo-xylosaccharide substrates to the active site is mainly mediated by hydrogen bonds and stacking interactions. Each region of the enzyme that accommodates a sugar xylose unit (or other sugar moieties) of the oligomeric sugar substrate is commonly considered to be a binding subsite. The general scheme adopted for sugar-binding subsites in xylanases and other glycosyl hydrolases follows the $-n$ to $+n$ nomenclature, where $-n$ represents the nonreducing end (glycon region) and $+n$ represents the reducing end (aglycon region). According to this scheme, catalytic cleavage takes place between the -1 and $+1$ subsites (Davies *et al.*, 1997; Pell, Szabo *et al.*, 2004). Family 10 xylanases usually accommodate between four and seven substrate-binding subsites (Pell, Szabo *et al.*, 2004) and most commonly five or six.

The thermophilic aerobic bacterium *G. stearothermophilus* T-6 possesses an efficient and complete xylan-degradation system. The system is based on a single extracellular xylanase that breaks down the extracellular polymer into short substituted xylosaccharides. These shortened modified oligosaccharides can readily enter the cell by virtue of specific ABC sugar transporters (Shulami *et al.*, 2007). Inside the cell, other cleaving enzymes such as α-arabinofuranosidase and α-glucuronidase remove the side substitutions, resulting in short unmodified (linear) xylosaccharides. The final hydrolysis of these xylosaccharides requires the action of an intracellular xylanase and exo-xylosidases. This physiological strategy allows the bacterium to successfully compete in nature with other plant cell-wall degrading microorganisms. Interestingly, in *G. stearothermophilus* both the extracellular xylanase and the intracellular xylanases belong to family 10 and share 41% sequence identity and 57% sequence similarity. Although both enzymes hydrolyze the same chemical bond, their substrate specificity towards xylan and its fragments is significantly different. Therefore, these two enzymes represent an excellent model to investigate substrate specificities among family 10 xylanases.

In the framework of this study, we recently crystallized the extracellular xylanase (XT6) isolated from *G. stearothermophilus* T-6 in two crystal forms (hexagonal; Teplitsky *et al.*, 1998; monoclinic; Bar *et al.*, 2004). The three-dimensional structure of the hexagonal form was then determined at 2.4 Å resolution, demonstrating its overall fold, the location of its active site and the main residues involved in its catalytic mechanism (Teplitsky *et al.*, 2004). In the present report, we describe the crystallographic analysis and the resulting three-dimensional structure of the corresponding intracellular xylanase (IXT6) of the same bacterium at 1.45 Å resolution. This structure is then used for comparison with XT6 and other xylanases of family GH10 in order to draw more general conclusions regarding the structure–specificity relationships of enzymes within this family.

2. Materials and methods

2.1. IXT6 crystallization

Expression, purification and preliminary crystallographic analysis of the intracellular β-1,4-D-xylanase IXT6 were carried out as described previously (Teplitsky *et al.*, 2000). The crystals initially obtained using the published procedure did not diffract X-rays beyond 2 Å resolution. Further refinement of these conditions resulted in improved crystals which often diffracted to beyond 1.5 Å resolution. These improved crystals

Table 1
Representative parameters from the crystallographic data measurement of IXT6.

Values in parentheses are for the outer diffraction shell.

Beamline	NSLS X26C
Wavelength (Å)	1.100
Space group	*C*2
Unit-cell parameters	
a (Å)	169.48
b (Å)	80.58
c (Å)	79.05
β (°)	91.89
No. of reflections	
Total	1266239
Unique	171850
Redundancy	7.0 (5.5)
$\langle I\rangle/\langle\sigma(I)\rangle$	18.1 (1.5)
Mosaicity	0.46
Resolution range (Å)	38.0–1.45 (1.48–1.45)
Completeness (%)	91.6 (52.3)
$R_{\rm sym}$ (%)	5.80 (44.4)

were obtained using the hanging-drop vapour-diffusion technique at room temperature. The drops (10 µl) were formed by mixing 5 µl protein solution (6 mg ml^{-1}) in 50 m*M* Tris–HCl buffer pH 7.0, 100 m*M* NaCl and 0.02% NaN_3 with an equal volume of reservoir solution containing 0.1 *M* sodium cacodylate pH 6.5 and 1.9 *M* sodium acetate (citric acid was used as precipitant in the original conditions). These drops were equilibrated against 1.0 ml reservoir solution at a constant temperature of about 293 K. Well formed crystals (usually grown in a polyhedron 'diamond' shape with distinct faces and sharp edges) appeared after 3–4 d and grew to full size in about two weeks. Typical dimensions of the fully grown crystals were 0.3 × 0.3 × 0.5 mm.

2.2. Crystallographic data collection

The fully grown crystals were soaked (for about 20–30 s) in a cryosolution (70% reservoir, 30% glycerol) prior to direct flash-cooling in a nitrogen-gas cold stream. Diffraction data collection was carried out at 95 K using synchrotron radiation (λ = 1.10 Å) and a CCD area detector (Q4, ADSC, USA) at the X26C beamline, National Synchrotron Light Source (NSLS) facility, Brookhaven National Laboratory, NY, USA. A lower resolution data set (a total of 215 frames) was initially collected to 2.0 Å (crystal-to-detector distance of 200 mm) using 1.0° oscillation frames and 5 s exposures. Since no significant radiation damage was observed, the same crystal was then used for high-resolution data measurement to 1.45 Å resolution (crystal-to-detector distance of 110 mm, 1° oscillations, 10 s exposure, 375 frames). Data were processed, integrated and scaled with *DENZO* and *SCALEPACK* (Otwinowski, 1993; Otwinowski & Minor, 1997). These data indicated that the crystals belong to a *C*-centred monoclinic crystal system (space group *C*2), with crystallographic unit-cell parameters *a* = 169.48, *b* = 80.58, *c* = 79.05 Å, β = 91.89° and with two IXT6 molecules in the crystallographic asymmetric unit. Representative data-collection parameters of the combined data set are listed in Table 1.

2.3. Structure determination

The structure of native IXT6 was solved by molecular replacement (MR; Navaza & Saludjian, 1997) using the structures of XT6 in its trigonal form (2.4 Å resolution; PDB code 1hiz; Teplitsky *et al.*, 2004) and in its monoclinic form

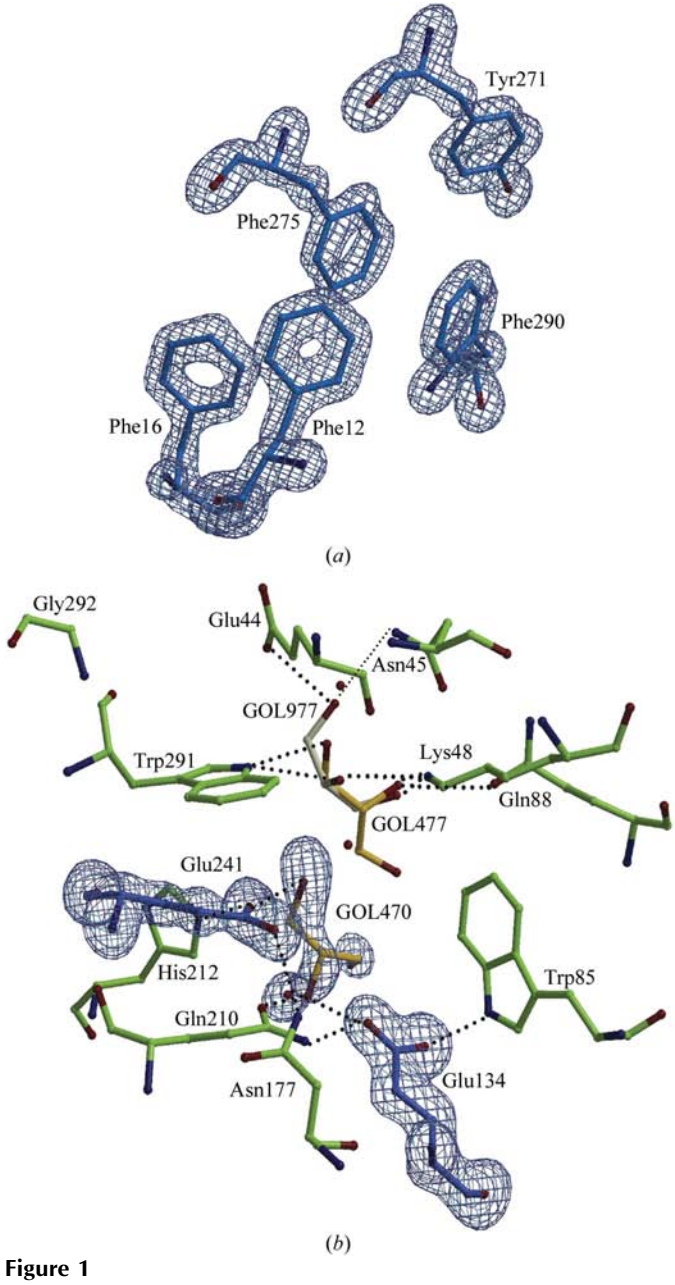


Figure 1
Representative sections of the electron-density map (OMIT map at the final stage of the refinement, contour level of 3.0σ, cyan) demonstrating the quality and reliability of the map interpretation of the current structure of IXT6 at 1.45 Å resolution. Superimposed on the map are the corresponding regions of the final model of IXT6 (conventional atom colours). (*a*) The region around residues 12, 16, 271, 275 and 290 of molecule *A*, showing the clear density and straightforward model building of typical aromatic side chains. (*b*) The region of the two catalytic residues, Glu134 and Glu241, together with the main glycerol molecule (GOL470) found to be bound in the active site. The second (disordered) glycerol molecule (GOL477/977) is also included. Hydrogen bonds are indicated by dotted lines.

Table 2
Representative statistics from the crystallographic refinement of IXT6.

Model refinement	
Data resolution range (Å)	38.0–1.45 (1.50–1.45)
Data cutoff	$\|F\| > 0$
R factor† (%)	15.0
No. of reflections	156685
R_{free}† (%)	19.0
No. of reflections	6741
Refined model	
No. of residues‡	658
No. of protein atoms	5604
No. of solvent atoms	765
Average B factor (Å^2)	30.69
R.m.s. deviation	
Bond lengths (Å)	0.012
Bond angles (°)	2.23
Solvent content (%)	64
No. of molecules per ASU	2
Estimated coordinate error (Å)	
Luzzati	0.08
σ_A	0.10
Ramachandran plot	
Residues in most favoured regions (%)	91.1
Residues in additionally allowed region (%)	8.9

† $R = \sum ||F_o| - |F_c|| / \sum |F_o|$. A random subset (about 5%) of the data was used for the calculation of R_{free}. ‡ 330 residues in molecule A; 328 residues in molecule B; 5604 atoms in the total content of the asymmetric unit (molecule A + molecule B).

(1.45 Å resolution; PDB code 1r85; Bar *et al.*, unpublished results) as reference search models. The 1r85 model of XT6 turned out to give significantly better MR results. The best solution obtained using this reference model gave a figure of merit (FOM) of 48%, confirming that there are two IXT6 monomers in the crystallographic asymmetric unit. After density modification, the FOM for this MR solution improved to 74%. The corresponding initial model of IXT6 gave an overall crystallographic R factor of 50%, improving to an R factor of 43% after the first cycle of rigid-body refinement with the program *CNS* (Brünger *et al.*, 1998). The full model of IXT6 was built into the electron density using the program *O* (Jones *et al.*, 1991).

2.4. Model building and refinement

Refinement of the initial model of IXT6 was performed using *CNS*, employing the maximum-likelihood amplitude target procedure (Brünger *et al.*, 1998). This model, which consisted of two independent IXT6 molecules in the asymmetric unit, was subjected to rigid-body refinement, simulated annealing and iterative refinement cycles of positional parameters and temperature factors. Each refinement round was followed by manual fitting and rebuilding of the protein model using the program *O*. Missing protein residues were built step by step and adjusted into the $(2F_o - F_c)$ and $(F_o - F_c)$ electron-density difference maps. This procedure greatly improved the quality of the electron-density maps and allowed a gradual reconstruction of almost all of the protein residues. Water molecules were assigned to peaks in the $(F_o - F_c)$ difference electron-density maps at a contour level greater than 3.4σ that were also at a suitable distance and in a suitable orientation to form a hydrogen bond with a potential partner. This procedure resulted in 680 such water molecules that were included in the crystallographic model in the final cycle. Once most of the residues of the two independent IXT6 molecules of the model had been built, the refinement procedure was switched to the program *SHELX*-97 (Sheldrick & Schneider, 1997), which allowed fully refined anisotropic temperature factors for most of the individual atoms of the model. This refinement converged to a final R factor of 14.98% and a final R_{free} of 19.01% (for about 5% of the data; Brünger, 1992). Representative sections of the final electron-density map are shown in Fig. 1, demonstrating its excellent quality and its unequivocal interpretation. Representative parameters for the refinement and the final structural model of IXT6 are listed in Table 2.

2.5. Content of the final crystallographic model

The final model comprises most of the amino-acid residues of the two independent IXT6 molecules (labelled *A* and *B*) in the asymmetric unit. A number of N-terminal residues of both molecules (Met1*A*, Met1*B*, Asn2*B* and Ser3*B*; *A* and *B* correspond to the specific monomer of the crystallographic model) were not included in the final model as the corresponding experimental electron density was too weak to define their positions with confidence. Relatively weak electron density (indicating conformational flexibility and/or crystallographic disorder) was also observed for Asn2*A*, Arg75*A*, Gln98*A*, Glu139*A*, Glu185*A*, Glu262*A*, Arg75*B*, Asp138*B* and Glu142*B*, for which the atoms of the main chain could be modelled with confidence but not those of the side chains. As a result, Ala or Gly residues were modelled into the electron density at these positions. All of these residues are located on external loops of the protein which are exposed to solvent, accounting, at least in part, for their relative conformational 'flexibility'. Interestingly, most of these flexible residues are part of the substrate-binding site, suggesting their potential involvement in the opening and closing of the active site for substrate binding and product release. As in other GH10 xylanases, three *cis*-peptide bonds were found along the polypeptide chain. These are located between residues His81 and Thr82, Arg217 and Pro218, and between Phe303 and Pro304. A total of 14 amino-acid side chains (103*A*, 116*A*, 120*A*, 186*A*, 204*A*, 228*A*, 236*A*, 256*A*, 263*A*, 274*A*, 284*A*, 312*A*, 327*A* and 330*A*) were modelled in two alternate conformations in the final model of IXT6 molecule *A*. Similarly, 14 side chains (73*B*, 116*B*, 118*B*, 148*B*, 159*B*, 167*B*, 170*B*, 182*B*, 224*B*, 261*B*, 263*B*, 286*B*, 287*B* and 320*B*) were modelled in two alternate conformations in the final model of molecule *B*.

The calculated solvent content of the unit cell is about 64%, based on the unit-cell volume and its confirmed protein content. Of this crystallographic solvent, only 680 water molecules were identified unequivocally and refined in the final model, as described above. Additional positive electron densities in the active site and close to the protein surface were assigned as 13 independent glycerol molecules, based on their difference density and their contacts with neighbouring

functional groups. These molecules most likely originated from the cryoprotectant solution (which contained 30% glycerol) in which the crystal was soaked prior to data measurement. Two sodium ions (one per protein molecule) were also assigned in the final model on the basis of their electron density, their coordination geometry (close to trigonal bipyramidal) and their typical metal–ligand distances (Harding, 2002). The final content of the crystallographic asymmetric unit refined in the present work includes 5604 (non-H) protein atoms and 765 (non-H) nonprotein atoms (a total of 6369 non-H atoms per asymmetric unit).

2.6. Quality of the final model

As mentioned above, the high quality of the current structure is demonstrated by its clear and easily interpretable electron-density maps (Fig. 1). These maps allowed reliable conformational orientation of most of the side chains (Fig. 1*a*), as well as the bound glycerol molecules (Fig. 1*b*). The final structure of the protein was also examined using the program *PROCHECK* (Laskowski *et al.*, 1993), allowing a general validation of its main structural parameters. This analysis of the conformation of the polypeptide chain (Ramachandran *et al.*, 1963) showed that all of the residues fall into the accepted regions of the Ramachandran plot. More specifically, 91.1% of the nonglycine and nonproline residues in the asymmetric unit are located in 'most favoured' regions and 8.9% are located in 'additionally allowed' regions. None of the protein residues are located in 'disallowed' regions.

The protein molecules conform closely to standard bond lengths and angles, as defined by Engh & Huber (1991), with relatively small root-mean-square (r.m.s.) deviations of 0.012 Å and 2.23°, respectively. In the final crystallographic model, the average overall *B*-factor for main-chain atoms is 27.13 Å^2, the average *B*-factor for side-chain atoms is 29.81 Å^2 and the corresponding value for all atoms is 28.50 Å^2, indicating normal overall temperature factors and a reasonable interpretation of the observed electron density. Based on the resolution (1.45 Å) and the final *R* factor (15.0%), the average experimental error in the coordinates of the final model is about 0.1 Å (±0.08 Å according to the Luzzati error estimation; Luzzati, 1952), permitting a meaningful and reliable analysis of the interactions and geometries involved in the structure presented here. The final structure of IXT6 (protein alone; molecule *A*) is schematically shown in Fig. 2. The final coordinates of the full structure have been deposited in the PDB under the identification codes 1n82 and 2q8x (where 2q8x is the latest and more refined model of the two entries).

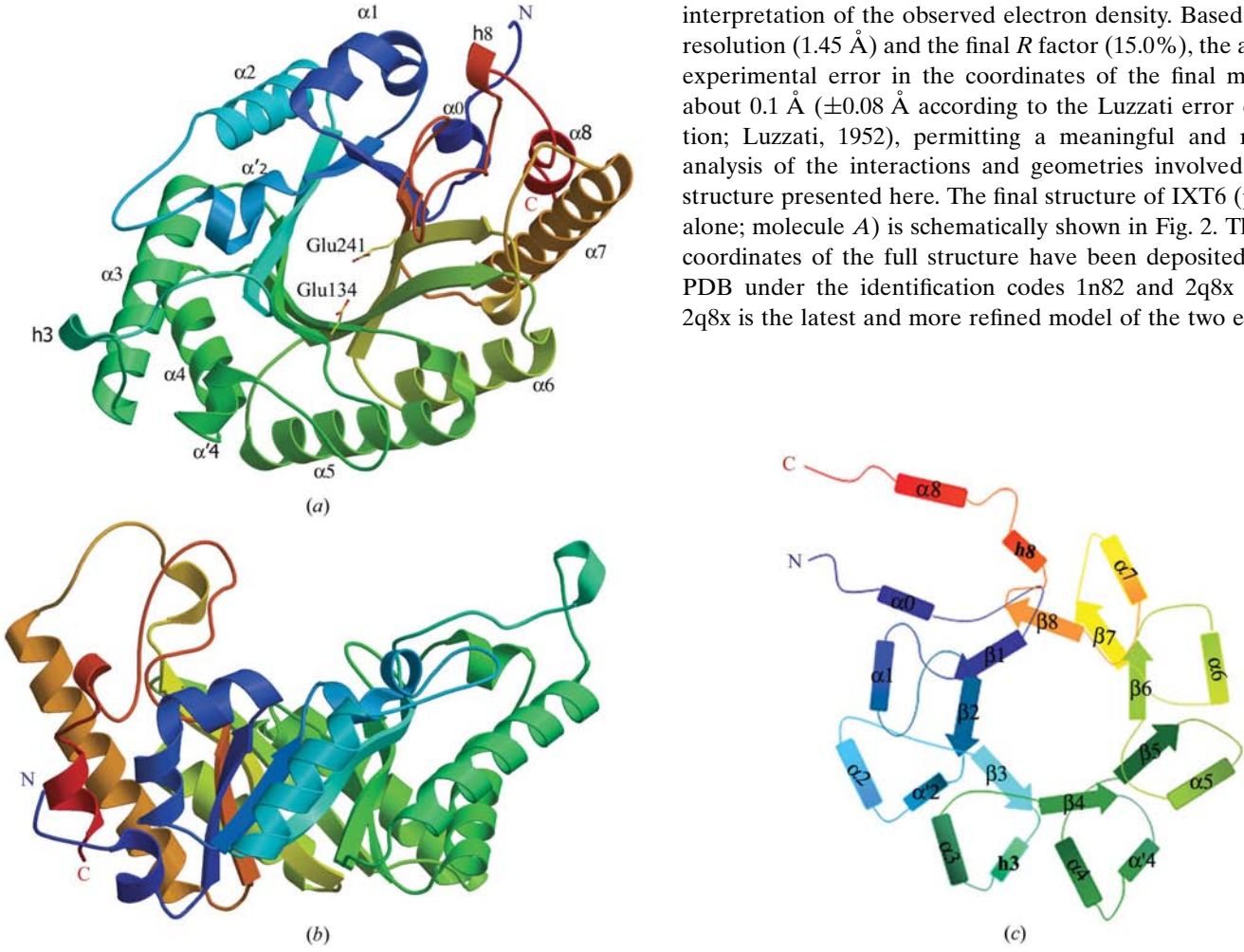


Figure 2
The overall structure of IXT6 (molecule *A*) resulting from the current crystallographic analysis at 1.45 Å resolution. The polypeptide chain of the enzyme is 'rainbow' colour-coded so that it starts (amino-terminus; N) in dark blue and then gradually changes to blue, light blue, blue–green, light green, green, green–yellow, yellow, yellow–orange, orange, orange–red and red (carboxy-terminus; C), demonstrating the overall structure of the enzyme and its secondary-structure elements. (*a*) A ribbon representation (top view), showing the general TIM-barrel fold of the molecule typical of GH10 xylanases. The two catalytic residues, Glu134 and Glu241, are shown in the central cavity. (*b*) A ribbon representation (side view) demonstrating the 'salad-bowl' overall shape of the enzyme. (*c*) A schematic topology diagram of IXT6. 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.

2.7. Calculations and figure preparation

The matrices for the superposition of the xylanase structures were calculated using a least-squares distance-minimization algorithm (*LSQ*, implemented within the program *O*), using the active-site C^{α} atoms as the guide coordinates. Fig. 1 was prepared using a combination of the programs *MOLSCRIPT* (Kraulis, 1991) and *BOBSCRIPT* (Esnouf, 1997). Figs. 2, 3, 4(*b*) and 5 were prepared using the program *MOLSCRIPT* (Kraulis, 1991). Fig. 4(*a*) was prepared using the program *LIGPLOT* (Wallace *et al.*, 1995). Figs. 4(*c*), 4(*d*) and 7 were prepared using the program *SPOCK* (Christopher, 1998). Fig. 5(*d*) was prepared using the program *SUPERPOSE* from *CCP*4*i* (Krissinel & Henrick, 2004). Fig. 6 was calculated and prepared using *ClustalW* (Thompson *et al.*, 1994) together with the *ESPript* server (Gouet *et al.*, 1999). Fig. 8 was prepared using the program *CHIMERA* (Pettersen *et al.*, 2004).

3. Results and discussion

3.1. Structure determination of IXT6

The general crystallization conditions and preliminary crystallographic characterization of WT IXT6 have been described in detail previously (Teplitsky *et al.*, 2000). Only relatively small changes to the crystallization procedure were needed to improve the quality of these crystals and they enabled diffraction data collection to a resolution limit of 1.45 Å for the present crystals. These diffraction data were used for structural analysis of the intracellular IXT6 by molecular replacement (MR) using the known structure of the extracellular xylanase from the same bacterium, XT6, as the reference model. The MR solution was clear and unequivocal, indicating that there are two molecules of the protein in the crystallographic asymmetric unit. The calculated MR phases proved to be of sufficient quality to produce interpretable electron-density maps of WT IXT6, which were further improved by density-modification (solvent-flattening) procedures. Even at this stage the resulting map was clean and clear enough to show the main features of the enzyme, including secondary-structure elements, solvent molecules and local hydrogen bonds, and the solvent channels between them. After several rounds of refinement and model adjustment of the initial structure, a new model for WT IXT6 was obtained which was then fully refined against the current 1.45 Å diffraction data set to obtain the final structure reported and discussed here (Fig. 2). The relatively high quality of the present structure is demonstrated, among other things, by the atomic resolution of most of the amino-acid side chains (Fig. 1*a*) and the clear identification of the bound solvent molecules (Fig. 1*b*), the relatively good final crystallographic *R* factor and the normal distribution of side-chain conformations in the Ramachandran plot (Table 2). This structure is thus reliable enough to analyze not only the general structure of IXT6 alone, but also the fine details of its active site and the features that distinguish it from XT6 and other xylanases.

3.2. Content of the crystallographic asymmetric unit

There are two molecules of IXT6 in the asymmetric unit of the current crystal structure, with a general noncrystallographic twofold axis relating them to each other (Fig. 3*a*). The two protein molecules are practically identical and superimposition of them (not shown here) demonstrates almost complete overlap. The discussion below thus relates equally to the two independent IXT6 molecules analyzed in the present structure, although specific geometric features, hydrogen bonds, angles and distances are given (if not specifically indicated) only for molecule *A*, which is less disordered and for which the experimental electron density is relatively more complete. As in similar cases of crystal structures with more than one molecule in the asymmetric unit, the overall structural identity of the two independent molecules in the current structure gives additional support to the biological significance of the crystal structure, since the two independent molecules differ in their crystallographic packing and intermolecular contacts. It should be noted, however, that the two molecules

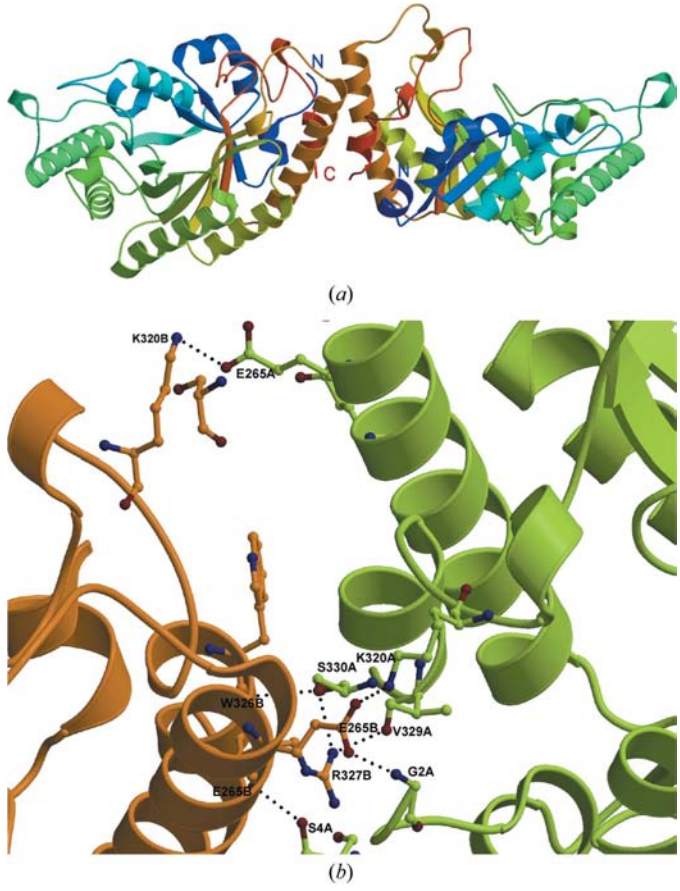


Figure 3
The intermolecular arrangement of the two independent IXT6 molecules in the crystallographic asymmetric unit. (*a*) The general orientation of the two molecules relative to each other (left, molecule *A*; right, molecule *B*). The two molecules are related by an approximate noncrystallographic C_2 axis. (*b*) A close-up view of the contacts between the two molecules, demonstrating the specific hydrogen bonds that connect the two IXT6 monomers (green, molecule *A*, residues labelled *XnnnA*; yellow, molecule *B*, residues labelled *XnnnB*).

show some small, but probably insignificant, conformational differences (r.m.s.d. of 0.35 Å for the C^{α} atoms of 328 residues), especially in the disordered sections and near the residues observed in alternate conformations.

Interestingly, the two IXT6 molecules of the asymmetric unit form an actual dimer in the crystal that is bonded *via* several intermolecular hydrogen bonds (Fig. 3*b*). These bonds are formed between residues located in the C-terminal region of the two IXT6 molecules, all of which are not conserved. Specifically, bonds are formed between Glu26 OE2 of molecule *A* and Lys320 NZ of molecule *B* and similarly between Glu265*A* OE2···Lys320*B* NZ, Lys320*A* NZ···Glu265*B* OE2, Val329A O···Arg327*B* NH1, Ser330*A* OG···Trp326*B* O and Ser330*A* OG···Arg327*B* NH1. Based on gel filtration, IXT6 is present as a monomer in solution (results not shown). Taken together with the fact that the interacting residues are not conserved, it is likely that the 'dimerization' observed in the present crystal structure is not biologically significant and is probably formed only as a result of crystal-packing energetics.

3.3. The overall structure of IXT6

The overall structure of IXT6 (Fig. 2) corresponds to the classical eightfold α/β-barrel (TIM-barrel; Banner *et al.*, 1975). This fold is relatively common in enzyme structures (occurring in about 10% of all enzymes) and has been found for all GH10 xylanase structures reported to date (*e.g.* Derewenda *et al.*, 1994; Harris *et al.*, 1996; Schmidt *et al.*, 1998; Fujimoto *et al.*, 2000; Teplitsky *et al.*, 2004). As for most GH10 xylanases, the shape of the IXT6 α/β-barrel is elliptical in cross-section (Fig. 2*a*), with the major axis running between β-strands 1 and 5 of the $(\alpha/\beta)_8$ domain. As commonly observed, the top of the α/β-barrel is higher at the end points of the long elliptical cross-section axis and lower at the end points of the short elliptical axis, resulting in a bowl-shaped cavity (often referred as a 'salad-bowl'; Fig. 2*b*) which is about 15 × 10 Å wide and about 20 Å deep. The active site of the enzyme is located on the inner part of this cavity facing the solvent-accessible open region in its centre. In the present structure, the overall long elliptical axis of the barrel is about 60 Å, while the short elliptical axis of the barrel is about 40 Å. The maximum height at the side 'walls' of the barrel is about 40 Å and the shortest distance of the active site from the bottom of the internal cavity of the barrel is about 20 Å.

A schematic topology diagram of the current structure of IXT6 is shown in Fig. 2(*c*). The numbering of the α-helices (α_n) and the β-strands (β_n) was assigned in order to keep the classical arrangement of eight β-strands (β_1–β_8) at the internal core of the $(\alpha/\beta)_8$-fold and eight α-helices (α_1–α_8) at the external part of the $(\alpha/\beta)_8$-fold (Fig. 2*a*). The β-sheet strands of the TIM-barrel are accordingly constituted of residues 17–22 (β_1), 39–44 (β_2), 77–86 (β_3), 127–134 (β_4), 173–179 (β_5), 205–215 (β_6), 236–245 (β_7) and 284–291 (β_8). Similarly, the α-helices of the α/β-barrel are constituted of residues 23–36 (α_1), 61–74 (α_2), 103–123 (α_3), 147–154 (α_4), 183–200 (α_5), 219–232 (α_6), 260–281 (α_7) and 322–330 (α_8).

IXT6 has several additional helices that are part of the general α/β-barrel but are outside the basic TIM-barrel fold. α_0 is located at the N-terminal polypeptide (residues 7–12), just prior to β_1. h_3 (residues 90–95) and h_8 (residues 296–302) are two short 3_{10}-helices that are part of the two loops leading to the active-site core residues. In the polypeptide sequence of IXT6, these two helices are located right after β-strands β_3 and β_8, respectively. Two additional short α-helices, α_2' (49–53) and α_4' (156–169), are located just prior to α_2 and α_4, respectively (Fig. 2*c*).

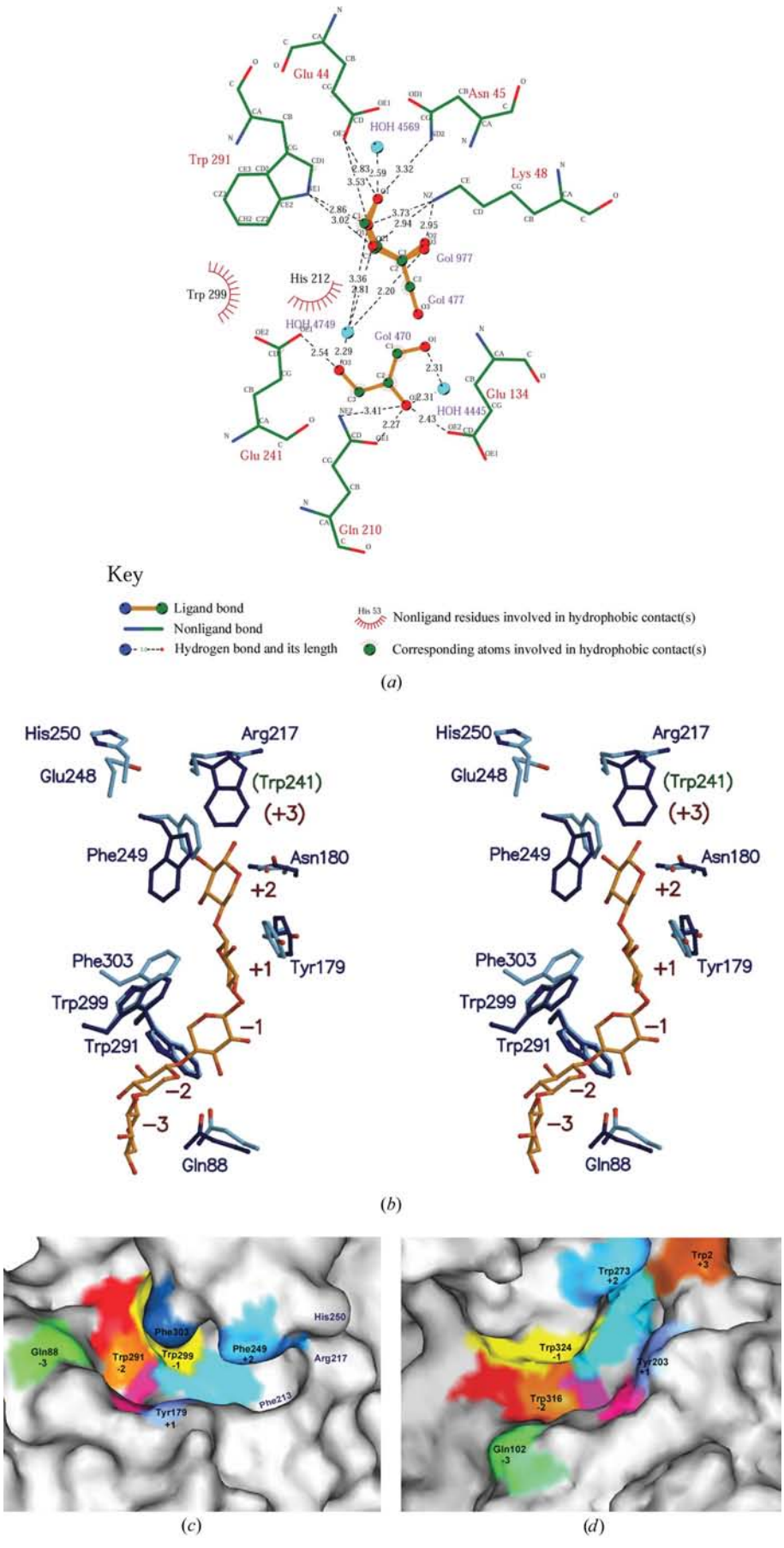
Hence, except for several small secondary-structure elements, IXT6 very closely follows the canonical TIM-barrel fold. As is common for GH10 enzymes, a relatively long and narrow groove is formed on the protein surface at the carboxy-terminal end of the internal β-barrel. This groove is exposed to the solvent and is lined by an array of aromatic and hydrophilic residues; its structure and dimensions can accommodate short oligoxylose substrates, as discussed below.

3.4. Special structural features

A total of 13 glycerol molecules (originating from the glycerol-rich cryoprotecting solvent) were introduced into the electron-density map and are part of the final crystallographic model of IXT6. Most of these molecules appear on the surface of the protein and are bound by at least one hydrogen bond to surface functional groups (charged or polar). Two of these glycerol molecules were found in the active site (in both of the IXT6 molecules of the asymmetric unit), bound to the two acidic catalytic residues (Glu134 and Glu241) in a similar fashion expected for a xylose substrate/product (see below). In general, all 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 that are likely to host water molecules in purely aqueous solution. We thus expect the structure of IXT6 to be practically identical in a glycerol-free (physiological) environment.

Three *cis*-peptide bonds are observed in the structure of IXT6, two of which are proline-related and located between residues 217–218 and 303–304. The third *cis* bond is located between His81 and Thr82 and seems to be highly conserved in GH10 xylanases. This bond, which is generally located in the active-site area, causes a bend in the polypeptide chain that brings His81 into a special position in the substrate-binding region. Indeed, based on several enzyme–substrate complexes (*e.g.* Schmidt *et al.*, 1998; Ducros *et al.*, 2000), the corresponding histidine seems to be part of the interaction of the enzyme with the bound substrate at subsite −1.

A nonprotein cation (probably sodium) was found to be bound to the protein (in both monomers in the asymmetric unit) between strand β_8 and the short polypeptide loop at the N-terminus (located between the α_0 helix and the β_1 strand). This cation is coordinated to a water molecule and four protein residues (Ile285 O, Arg282 O, Gln286 OE1 and Asp15 O) in a near-trigonal bipyramidal geometry. This sodium coordination seems to stabilize the local conformation of the TIM-barrel around strands β_1 and β_8 and, in particular,



to allow a better closure of the N-terminal and C-terminal parts of the barrel. Interestingly, no other examples of sodium-mediated barrel closure have been reported for GH10 xylanases, although some cases of calcium-mediated stabilization have been described (*e.g.* for xylanase A from *Pseudomonas fluorescens*; Harris *et al.*, 1996). In the case of the BSX xylanase from *Bacillus* sp. NG-27, an Mg^{2+} ion (probably from the crystallization solution) was found at the C-terminal end of the catalytic domain. The study of the effect of the Mg^{2+} on the activity of the enzyme revealed increased activity of BSX in a concentration-dependent manner. These results suggested that in the special case of BSX a metal ion is involved in the activity and stability of the enzyme (Manikandan *et al.*, 2006).

3.5. The active site

The active site of GH10 xylanases usually forms an open cleft surrounded by loops that are rich in aromatic residues. These residues are commonly involved in substrate binding. This is also the case in the active site of IXT6. The two catalytic residues, Glu134 and

Figure 4
The active site of IXT6 as determined in the present structure at 1.45 Å resolution. (*a*) A schematic diagram of the interactions around the active site, including those of the two bound glycerol molecules (GOL470 and GOL477/977) and the crystallographically determined water molecules (HOH). The protein bonds are shown in green and the bonds of the bound glycerol molecules are shown in orange. The remaining atoms have standard atomic colours. Dashed lines indicate hydrogen bonds or ionic interactions (distances are given in Å), while 'radiating' spheres indicate hydrophobic contacts. (*b*) Superposition of the active site of IXT6 (residues shown in cyan) on the active site of the XT6–pentaose complex (PDB code 1r87; protein residues, blue; bound sugar, orange). Residue numbers are given according to IXT6, except for Trp241 of XT6. Substrate-binding subsites are marked according to XT6. (*c*) Solvent-accessible surface representing the substrate-binding cleft of IXT6. The predicted binding subsites (−3 to +2) are coloured and labelled. The main subsite residues are marked and the two catalytic residues are shown in purple. (*d*) A similar representation of the XT6 binding cleft with corresponding subsite colours and labels (−3 to +3), including the main homologous binding residues.

Glu241, are located on opposite sides of the surface cleft (Fig. 2*a*), approximately a third of the way along the cleft. The distance between their functional carboxylate groups (5.54 Å) is consistent with the expected 'retaining' hydrolytic mechanism (Davies & Henrissat, 1995; White & Rose, 1997; Zechel & Withers, 1999). Based on sequence homology and the structures of homologous GH10 xylanases Glu241 is the nucleophile, while Glu134 is the acid–base catalyst (White *et al.*, 1994; Harris *et al.*, 1996; Schmidt *et al.*, 1998; Teplitsky *et al.*, 2004).

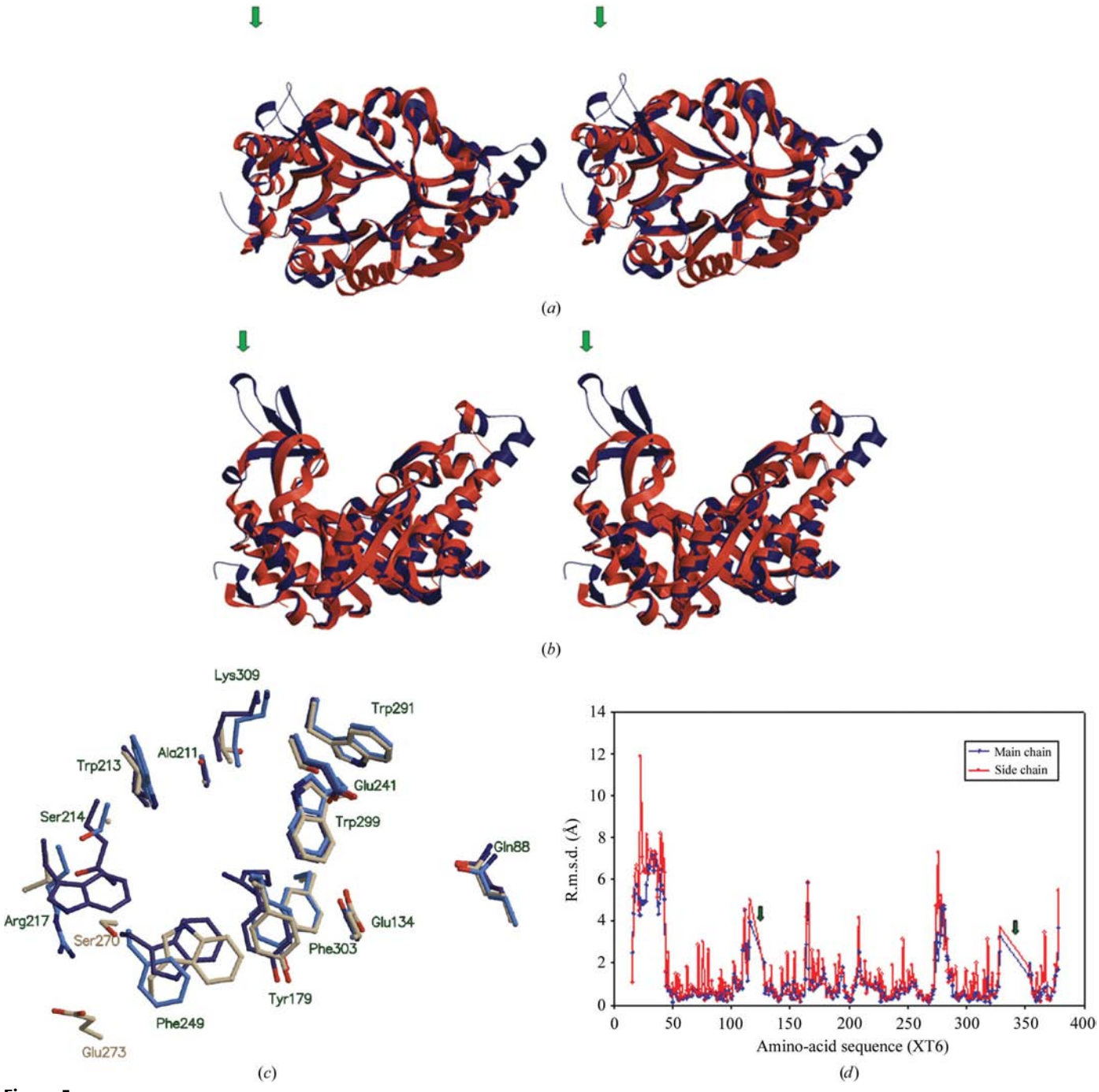


Figure 5
Structural superposition of the two xylanases of *G. stearothermophilus*, IXT6 and XT6. (*a*) Top stereoview of the overall structures of the two enzymes: IXT6 is shown in orange and XT6 in blue. (*b*) Side stereoview of the overall structures (same colour codes), demonstrating the 'salad-bowl' shape of the molecules. The unique subdomain of XT6 is marked. (*c*) Superposition of the active sites of the two enzymes together with the corresponding active site of XylC from *C. mixtus* (PDB code 1uqy), indicating the main residues that participate in substrate binding in the three enzymes. Amino-acid residues are numbered according to IXT6. IXT6 is shown in cyan, XT6 in blue and XylC in beige. Note the two extra charged/polar residues (Ser270, Glu273) at subsites +3 and +4 of the XylC active site (marked in beige). (*d*) A plot of the C^{α} r.m.s. deviation between the structures of IXT6 and XT6 as a function of the amino-acid residue number (according to the longer sequence of XT6). Green arrows indicate the two gaps in the IXT6 sequence correlated with the unique subdomain of XT6 (residues 330–350) and the additional loop at the opposite side (residues 116–128).

The catalytic nucleophile (Glu241) is located in a β-bulge on the β7 strand, while the acid–base catalyst (Glu134) is located in a β-bulge in the C-terminal part of the β4 strand. These positions are common for all glycoside hydrolases of the 4/7 superfamily, in which the two conserved glutamates are located near the carboxy-terminal ends of β-strands 4 and 7 (Jenkins *et al.*, 1995). Obviously, the exact position and orientation of these two catalytic carboxylate groups are vital for efficient enzymatic activity. These orientations are enabled by critical hydrogen bonds to neighbouring residues around the active site. The carboxylate group of the acid–base catalyst Glu134 is held in position by several such hydrogen bonds. One of its carboxylate group O atoms (OE1) is hydrogen bonded to Trp85 NE1, while the other oxygen (OE2) is hydrogen bonded to Gln210 NE2, as well as to O1 and O2 of the fully occupied glycerol molecule (GOL470) bound in the active site (Figs. 4*a* and 1*b*). Similarly, the O atoms (OE1 and OE2) of the carboxylate group of the catalytic nucleophile Glu241 are hydrogen bonded to His212 NE2, a conserved residue in GH10 xylanases that is considered to be responsible for maintaining the ionization state of the nucleophile (Roberge *et al.*, 1997). The Glu241 carboxylate O atoms are also hydrogen bonded to the fully bound glycerol molecule (O3) and to Asn177 ND2 (Figs. 4*a* and 1*b*).

As in other xylanases, the substrate-binding subsites of IXT6 are mainly defined by aromatic residues taking part in specific binding interactions with corresponding substrate sugar units. Here, it seems that Trp291 is part of subsite −2, Trp299 forms subsite −1, Tyr179 forms subsite +1 and Phe249 forms subsite +2. These subsites were assigned based on the superimposition of IXT6 and XT6 (Fig. 4*b*) and are schematically presented for both enzymes on the active-site surface in Figs. 4(*c*) and 4(*d*). These assignments are supported by thermodynamic binding measurements of xylo-saccharides in solution (Zolotnitsky *et al.*, 2004). According to such assignment of subsites, the main glycerol molecule of the active site (GOL470) is bound between subsites +1 and −1 (Fig. 4*a*), the exact position of the catalytic cleavage of 'normal' oligomeric sugar substrates. It is likely, therefore, that the protein interactions observed for this glycerol are also relevant for a real substrate, as should be validated by structural analyses of the corresponding enzyme–substrate and enzyme–analogue complexes. Such studies are currently under way.

3.6. Structural comparison between IXT6 and XT6

As mentioned above, *G. stearothermophilus* T-6 produces two homologous xylanases, XT6 and IXT6, both of which are utilized for xylan degradation. While XT6 is excreted by the bacterium and functions only extracellularly (on long, branched and decorated xylan substrates), IXT6 remains inside the cell and hydrolyzes short, linear and nondecorated oligoxylose substrates. These enzymes are of different sizes (XT6, 379 residues; IXT6, 331 residues) but share about 40% amino-acid sequence identity. With both structures available, it is now possible to compare the two enzymes in order to elucidate the differences in their substrate preferences. The two specific structures to be used for this comparison are the current structure of IXT6 at 1.45 Å resolution (PDB code 1n82) and the monoclinic structure of XT6, which was recently also determined at 1.45 Å resolution (PDB code 1r85; Bar *et al.*, unpublished results). Obviously, the equal resolutions of the two structures and the equivalent crystallographic procedures used to obtain them allow a more reliable comparison, with minimal experimental artifacts.

As expected, superposition of the two structures shows that the canonical TIM-barrel fold is practically identical in both structures (Fig. 5*a*). Nevertheless, the overall calculated r.m.s.d. between the polypeptide chains of the two structures is 1.08 Å (based on 317 common C^{α} atoms), indicating that there are conformational differences elsewhere in the structure. Most of these differences are located around the noncanonical features of both enzymes (Fig. 5*b*), namely the unique extra subdomain of XT6 (Teplitsky *et al.*, 2004), the small additional α-helices (α_{1a}, α_{3a}, α_{3b} and α_{7a}; XT6 notations) and β-strands (β_{3b} and β_{6b}) of XT6 and the small additional helices (h_3, h_8, α_2' and α_4') of IXT6 mentioned above. The unique subdomain of XT6 is probably the most pronounced difference between the two structures (Fig. 5*b*). This subdomain is located close to the C-terminal end ('top') of the β-barrel of XT6, between the classical eighth β-strand and the eighth α-helix (Teplitsky *et al.*, 2004). This added polypeptide chain of XT6 (constituted of residues 330–350) starts with a short α-helix (α_{8a}, seven amino acids), followed by a sequence of three short antiparallel β-strands (β_{8b}, β_{8c} and β_{8d}), each of which contains only two or three amino acids. These elements extend from the main part of the central TIM-barrel part of XT6 through a long bended loop (Fig. 5*b*). This distinct subdomain is a unique feature so far only found in XT6, which distinguishes XT6 not only from IXT6 but also from all other GH10 xylanases analyzed to date. The other extra loop of XT6, containing the α_{3a} and α_{3b} helices, is located on the other side of the central barrel, opposite to the unique subdomain of XT6 discussed above.

Interestingly, all those elements that differ between IXT6 and XT6 are located at the 'top' of both molecules, around the upper rim of the 'salad-bowl' (Fig. 5*b*). This area is relatively flexible, as expected from a region used for initial interactions with a large variety of incoming substrates. This is especially true for XT6, which should be able to work on branched and decorated xylan substrates, the accommodation of which is likely to be associated with large movements of the structural elements around the binding area. Another noticeable feature of the rims of the bowl, which is similar in both XT6 and IXT6, is that they differ in height from the bottom of the bowl. Such a difference is observed, for example, around the top of the α_3 helix, which is higher than the top of the α_8 helix at the opposite side of the rim. This, again, may bear functional significance, allowing substrates to approach the binding cleft (and probably to be directed into it) only from a specific side, correlated with the lower (and less abstracted) side of the top surface of the enzyme.

An analysis of the C^{α} deviation between the structures of IXT6 and XT6 as a function of the amino-acid sequence shows a good agreement with the points discussed above. Within the

α-helix or the β-sheet elements of the two xylanases a very high correlation is observed between the two structures (residue by residue), with very small r.m.s.d. values. These close correlations include the catalytic residues of both enzymes, 134 and 241 in IXT6 and 159 and 265 in XT6 (Fig. 5*d*). However, more significant differences between the

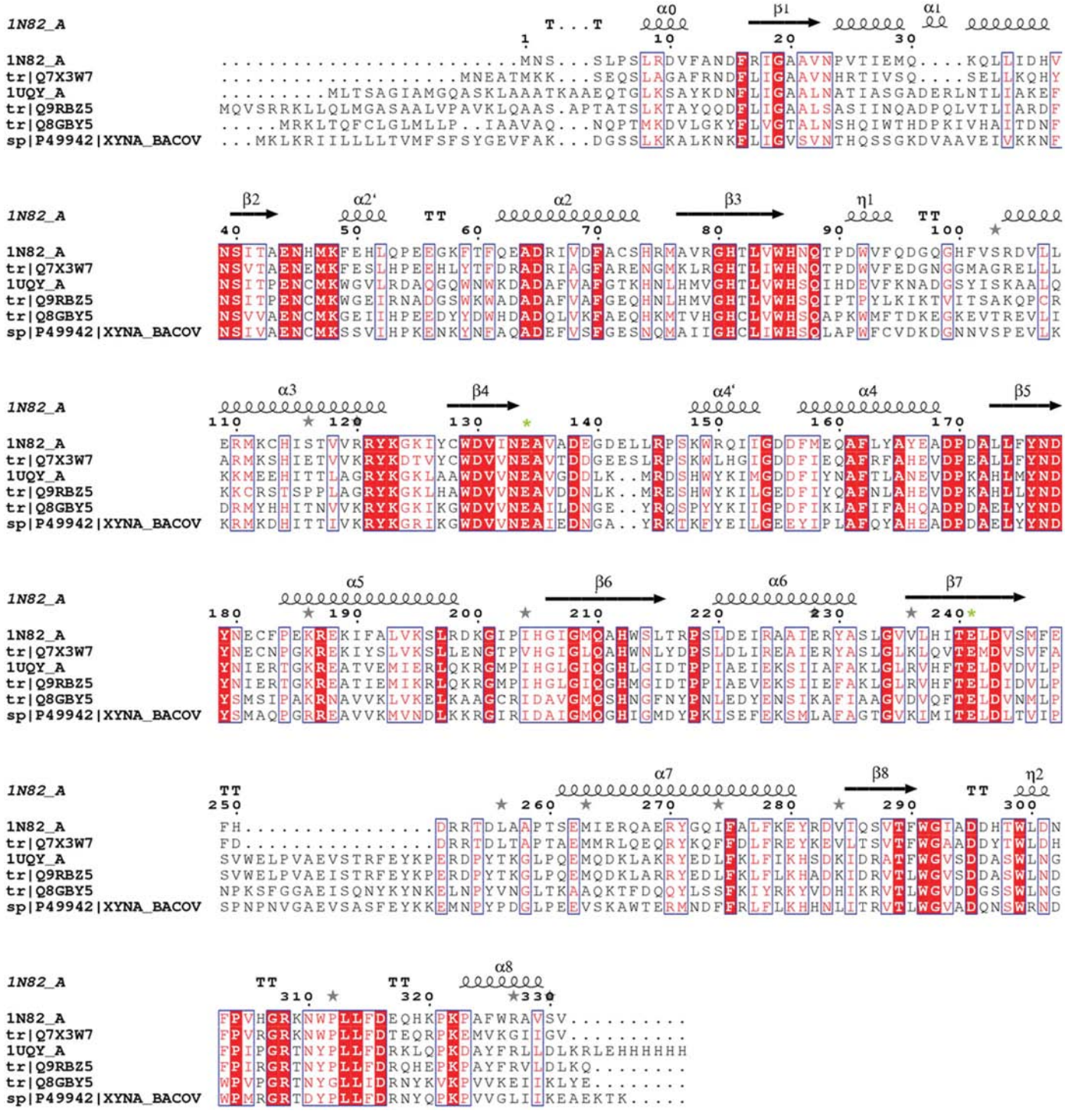


Figure 6
Amino-acid sequence alignment of IXT6 with other GH10 intracellular xylanases, as computed using *ClustalW* (Thompson *et al.*, 1994). The alignment includes the xylanases from *G. stearothermophilus* (IXT6; PDB code 1n82), *C. mixtus* (PDB code 1uqy), *C. japonicus* (XynG, Q9RBZ5), uncultured bacterium (Q7X3W7), *Bacteroides ovatus* (P49942) and *Prevotella bryantii* (Q8GBY5). Xylanases for which the three-dimensional structure is available are indicated by their PDB code. The other xylanases are indicated by their SwissProt sequence code. The two catalytic glutamate residues (E134 and E241 for IXT6) are marked above with an asterisk (*). Full red squares represent full identity of the residues among the species compared. Secondary-structure elements are marked above the relevant sections of the IXT6 sequences (coil, α-helix; arrow, β-strand) correlated with the scheme used in Fig. 2(*c*).

Table 3
Conservation of residues around the active sites of IXT6, XT6 and XylC.

IXT6	XT6	XylC
Conserved residues		
Gln88	Gln102	Gln111
Asn45	Asn59	Asn68
His212	His236	His233
Gln210	Gln234	Gln231
Tyr179	Tyr203	Tyr200
Glu44	Glu58	Glu67
Trp299	Trp324	Trp336
Trp85	Trp241	Trp108
Arg145	Arg168	Arg166
Nonconserved residues		
Trp213	Ile237	Leu234
Phe303	Arg328	Phe340
Arg217	Trp241	Thr238
Ala211	Ser235	Gly232
Phe249	Trp273	Trp272
Lys309	Lys356	Thr346
Ser214	Gln238	Gly235
Glu248	Gly272	Ser270
His250	Pro270	Val271

two structures are observed in the N- and C-terminal regions, the external loops and their edges (*e.g.* around residues 43, 165 and 208 of XT6), as well as in the region around the unique additional features of XT6 at the upper rim of the 'salad-bowl' at both sides (*e.g.* residues 329 and 275 on one side and residue 116 on the other). Side-chain deviations between the two structures mostly arise from alternative conformations (*e.g.* Lys72, Lys148 and Glu246) or from different amino-acid side chains (*e.g.* Arg23 *versus* Val, Gly165 *versus* Glu).

A close view into the substrate-binding clefts of IXT6 and XT6 shows significant differences in their overall shape and length (Figs. 4*c* and 4*d*). This cleft in IXT6 is shorter, is nearly blocked at the two ends, is narrower and its walls seem to be nearly perpendicular to the bottom of the cleft (Fig. 4*c*). The corresponding cleft in XT6 is much longer, is open at both ends, is generally wider and its walls seem to form a wider angle facing the outside solvent area (Fig. 4*d*). The narrower cleft of IXT6 is produced by several side chains that extend into the inner part of the cleft (*e.g.* Phe183, Trp213, Glu181 and Phe249), as well as several sections of the polypeptide main chain, which is displaced inwards into the cleft relative to XT6 (*e.g.* the main-chain atoms of residues 180–184 and 222–223). These differences in the size and shape of the binding clefts correlate well with the length and shape of the target substrates of the two enzymes, allowing the binding of the elongated and branched polymer of xylan in the open, longer and wider cleft of XT6, while limiting binding to only short and unbranched oligoxylose substrates in the blocked, shorter and narrower cleft of IXT6.

A definite subsite mapping of the binding cleft is usually possible only from high-resolution structures of the enzymes in complex with a series of oligosaccharide substrates. Such data are currently available only for XT6, the structure of which was recently determined together with substrates constituted of two to six xylose subunits (Teplitsky *et al.*, unpublished results). These data clearly demonstrate that there are six specific binding subsites in the binding cleft of XT6 (Figs. 4*b* and 4*d*). These subsites are designated as −3 to +3 from the nonreducing end to the reducing end of the sugar substrate, where substrate cleavage takes place between −1 and +1, as discussed above. For IXT6 such data are not available as yet, but from a structural overlap of the two structures (Fig. 4*b*) it seems that its binding cleft is made up of only five specific subsites, designated accordingly from −3 to +2 (Fig. 4*c*). Moreover, based on the structure of XT6 complexed with a pentaose substrate (PDB code 1r87), a few specific residues of IXT6 could be associated with the particular subsites identified above. These IXT6 amino-acid residues include Gln88 for subsite −3, Trp291 for subsite −2, Trp299 for subsite −1, Tyr179 for subsite +1 and Phe249 for subsite +2. The corresponding residues in XT6 are Gln102 (−3), Trp316 (−2), Trp324 (−1), Tyr203 (+1), Trp273 (+2) and Trp241 (+3) (Table 3 and Figs. 4*b*, 4*c*, 4*d* and 5*c*). Aromatic residues are usually ascribed to these subsites owing to their potentially important contribution to substrate binding *via* hydrophobic and aromatic stacking interactions. It was shown, however, that hydrogen bonds (contributed by non-aromatic residues) may also play a significant role in substrate binding. This was demonstrated recently for the intracellular xylanase from *C. mixtus* (XylC; PDB code 1uqy), in which the main residues in subsites +3 and +4 were found to be Glu273 and Ser270, respectively (Pell, Taylor *et al.*, 2004), yet the active site is otherwise very similar to both XT6 and IXT6 (Table 3, Fig. 5*c*).

Using isothermal titration calorimetry (ITC), Zolotnitsky *et al.* (2004) recently demonstrated that the preference of substrate binding in the active site of xylanases is based mainly on hydrophobic interactions and especially the stacking interactions of aromatic side chains with the xylose subunits of the substrate. Moreover, it was shown that ITC results can be directly correlated with the number of binding subsites and that this information can be used for the identification of the specific aromatic residues that are most critical for these interactions. According to these experiments, the binding differences between XT6 and IXT6 are located at the aglycon side of the substrate, correlated with subsites +2 and +3 of the binding clefts. For XT6 the main interacting residue in the +2 subsite is Trp273, while the corresponding residue in IXT6 is probably Phe249 (Table 3). IXT6 has an additional hydrophobic interaction at the +2 subsite created by Phe303 (homologous to Arg328 of XT6), which seems to be unique to xylanases that hydrolyze small substrate (see below). Another difference around subsite +2 is that Gln238 of XT6 is replaced by the smaller Ser214 residue in IXT6, accounting for overall tighter substrate binding at this subsite in IXT6 compared with XT6.

In contrast to the +2 subsite, the main interacting residue in the +3 subsite of XT6 is Trp241, which is replaced by a non-aromatic residue Arg217 in IXT6. In agreement, IXT6 appears to lack a specific substrate-binding subsite at this position. Indeed, at the edge of the aglycon side of the IXT6 binding cleft there are several residues such as Trp213, Glu248, His250 and Arg217, that seem to close the narrow cleft of the active site (Fig. 4*c*). XT6 has an isoleucine residue

(Ile237 in place of Trp213 in IXT6) at this position, leading to a more open cleft and allowing the possibility of binding longer substrates. These structural observations are consistent with the overall active-site shape of IXT6, which seems to be one of the shortest and narrowest among family 10 xylanases for which three-dimensional structures are available. This feature of IXT6 is graphically demonstrated in the schematic topologies of the active-site clefts of IXT6 and XT6, which are illustrated in Figs. 4(*c*) and 4(*d*), respectively.

3.7. A general comparison of IXT6 with GH10 xylanases

Several high-resolution structures of family 10 xylanases have been reported in the last 12 y, including that of xylanase A from *Streptomyces lividans* (Xyl10A; Derewenda *et al.*, 1994), the xylanase from *Cellulomas fimi* (CEX; White *et al.*, 1994), the xylanase from *Pseudomanas fluorescens* (XylA; Harris *et al.*, 1996), the xylanase from *Penicillium simplicissimum* (XlnA; Schmidt *et al.*, 1998) and the extracellular xylanase from *G. stearothermophilus* (XT6; Teplitsky *et al.*, 2004). All these structures present the classical $(\alpha/\beta)_8$ catalytic domain (Fig. 5*a*). Only relatively small variations are observed among the various structures, especially in the inner β-strands, and these are usually located around loops at the outer part of the eightfold α/β-barrel (Fig. 5*b*). In the current IXT6 structure these variations are represented by several small extra helices (h_3, h_8, α_2' and α_4'), all of which are located on the polypeptide chain slightly before the corresponding main helices of the barrel.

These loops and short helices are more flexible than the rest of the barrel and probably play a role in the orientation of, and transient interactions with, incoming substrates. As they differ in the various GH10 xylanases, these elements are likely to be involved in substrate binding and in determining substrate specificities (Schmidt *et al.*, 1998; Canals *et al.*, 2003), as should be verified when enzyme–substrate complexes of IXT6 become available.

3.8. Structural characteristics of intracellular xylanases

Only relatively few periplasmic or intracellular xylanases have been characterized and reported in the literature to date. This small subgroup includes those xylanases isolated from *Prevotella bryantii* B14 (XynA; Flint *et al.*, 1997; Miyazaki *et al.*, 1997), *Cellvibrio mixtus* (XylC; Fontes *et al.*, 2000; Pell, Taylor *et al.*, 2004) and *C. japonicus* (XylG; Fontes *et al.*, 2000). It also includes an intracellular xylanase from an uncultured bacterium (SwissProt No. Q7X3W7), xylanase A from *Bacteriodes ovatus* (P49942) and a xylanase from *Aeromonas*

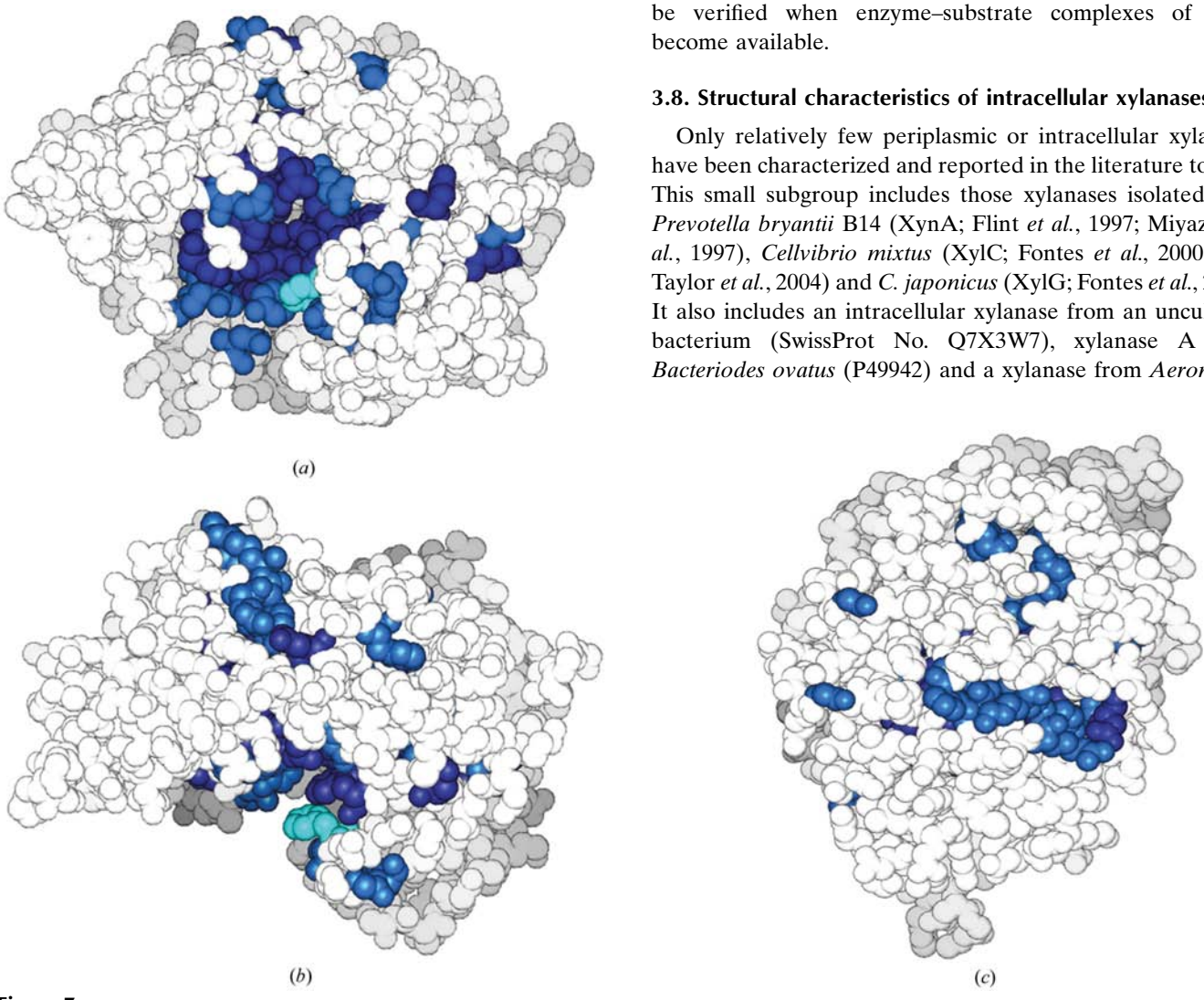


Figure 7
CPK model of IXT6 (produced with the program *SPOCK*). Conserved amino acids in family 10 xylanases are marked in dark blue, intracellular conserved amino acids are marked in light blue and the semi-conserved Phe303 (IXT6) is marked in turquoise. (*a*) Top view into the active site. (*b*) Side view of IXT6; about 90° horizontal rotation relative to (*a*). The active-site opening is seen at the bottom centre. (*c*) Back view of IXT6, demonstrating an additional patch of conserved residues. Since a number of these conserved residues are Arg and Lys residues, this patch may reflect a conserved (and potentially functional) positive electrostatic surface.

caviae (*punctata*) ME-1 (XynX; Usui *et al.*, 2003). Of these intracellular xylanases, only XylC from *C. mixtus* has been structurally analyzed and reported; its coordinates are available at the PDB with entry code 1uqy (Pell, Taylor *et al.*, 2004). This structure is thus the obvious reference for structural comparisons with the IXT6 structure reported here. As mentioned above, the active site of XylC is quite similar to that of IXT6 (and XT6), with the exception of its substrate-binding subsites +3 and +4 (Fig. 5*c*). Their similarity is also reflected in the relatively high number of homologous residues around their active sites, with a slightly higher correlation between the two intracellular xylanases IXT6 and XylC (Table 3).

The intracellular enzymes of family 10 have different polypeptide chain lengths, where the shortest is IXT6 (331 amino acids) and the longest are the xylanases from *C. japonicus* (XylG) and from *C. mixtus* (XylC) (378 amino acids each). Nevertheless, a sequence alignment of these intracellular enzymes (Fig. 6) reveals relatively more conserved sequences compared with the sequence alignment of the extracellular xylanases (not shown). Most of the conserved residues are located near the catalytic centre of the enzyme, as can be seen from their mapping on a CPK structural model of IXT6 (Fig. 7). Interestingly, those charged conserved residues (Arg, Lys and Asp) are usually hydrogen bonded to the catalytic glutamate residues (or to the substrate), while the aromatic conserved residues form hydrophobic interactions with the sugar rings of the bound substrate. Conserved areas of intracellular xylanases and especially 'positively charged patches' also appear on the other side (the 'back' surface) of the enzyme (Fig. 7*c*).

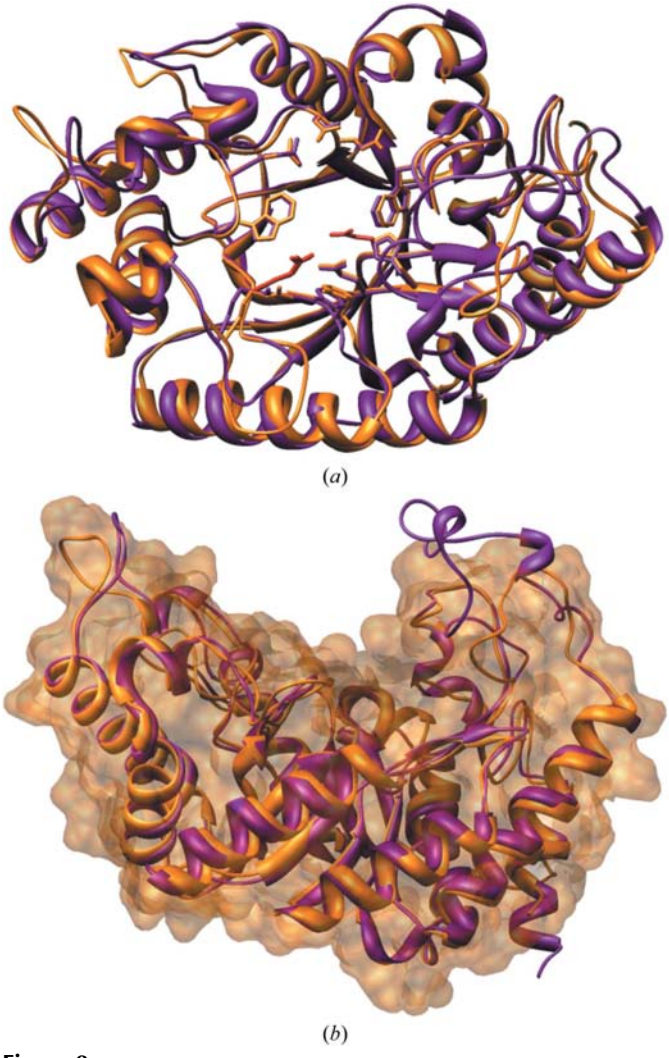


Figure 8
A superposition comparison of the current structure of IXT6 (gold) and the structure of the corresponding intracellular xylanase from *C. mixtus* (XylC; PDB code 1uqy; purple). (*a*) Top view of the overall structures of the two enzymes, demonstrating their very similar TIM-barrel central fold. Side chains involved in substrate binding are shown in ball-and-stick representation. (*b*) Side view of the overall structure of the two enzymes, demonstrating the differences in their outside loops. Also shown is a transparent surface model of IXT6 in the same orientation, providing a more realistic shape of the top and bottom faces of the enzyme.

In addition to the canonical $(\alpha/\beta)_8$-fold (Fig. 8*a*), the nonstandard secondary-structure elements for the XylC xylanase from *C. mixtus* include the extra helical segments α_1', α_2', α_3', α_7', α_7'' and α_8' (XylC notation; Fontes *et al.*, 2000; Pell, Taylor *et al.*, 2004). As expected, differences appear mainly around the higher edges of the bowl at the long axis of the ellipse, especially around the loops between β_7 and β_8, which act as a 'lid' to the active site, and between β_3 and β_4 at the other side of the active site (Fig. 8*b*). This different conformation of the loops is probably related to the topology of the cleft and the active site and may reflect the specificity of the enzyme for its specific substrates. Nevertheless, the exact role of these loops in specific enzyme–substrate interactions in the case of IXT6 should be determined by detailed and systematic structural analysis of complexes of this enzyme with a series of oligoxylose substrates and the corresponding substrate analogues. Such investigations are currently in progress.

This research was supported by grants from the GIF, the German–Israeli Foundation for Scientific Research and Development (No. 743-119 to GS and YS), and from the Israel Science Foundation (No 676/00 to GS and YS; No. 1006/05 to YS). Support was also provided by the Otto Meyerhof Center for Biotechnology at the Technion, established by the Minerva Foundation (Munich, Germany). VS was financially supported by a Joseph Klein fellowship and a Yenuka Yehuda scholarship from the Hebrew University and AT was supported by the Otto Schwartz fellowship of the Hebrew University. We thank the staff of the X26C beamline at the National Synchrotron Light Source facility (NSLS, BNL, NY, USA) for their support during X-ray synchrotron data measurement and processing. We also thank Gali Golan and Drs Rotem Gilboa and Vera Reiland for their helpful assistance in the structure analysis. YS holds the Erwin and Rosl Pollak Chair in Biotechnology.

References

Banner, D. W., Bloomer, A. C., Petsko, G. A., Philips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D. & Waley, S. G. (1975). *Nature (London)*, **255**, 609–614.

Bar, M., Golan, G., Nechama, M., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2004). *Acta Cryst.* D**60**, 545–549

Beg, Q. K., Kapoor, M., Mahajan, L. & Hoondal, G. S. (2001). *Appl. Microbiol. Biotechnol.* **56**, 326–338.

Brennan, Y. *et al.* (2004). *Appl. Environ. Microbiol.* **70**, 3609–3617.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Canals, A., Vega, M. C., Gomis-Rüth, F. X., Diaz, M., Santamaría, R. I. & Coll, M. (2003). *Acta Cryst.* D**59**, 1447–1453.

Christopher, J. A. (1998). *SPOCK: The Structural Properties Observation and Calculation Kit.* The Center for Macromolecular Design, Texas A&M University, College Station, TX, USA.

Collins, T., Gerday, C. & Feller, G. (2005). *FEMS Microbiol. Rev.* **29**, 3–23.

Coutinho, P. M. & Henrissat, B. (1999). *Carbohydrate-Active Enzymes Server (CAZY).* http://afmb.cnrs-mrs.fr/~cazy/CAZY/.

Davies, G. J. & Henrissat, B. (1995). *Structure*, **3**, 853–859.

Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem. J.* **321**, 557–559.

Derewenda, U., Swenson, L., Green, R., Wei, J., Morosoli, R., Shareck, F., Kluepfel, D. & Derewenda, Z. S. (1994). *J. Biol. Chem.* **269**, 20811–20814.

Ducros, V., Charnock, S. J., Derewenda, U., Derewenda, Z. S., Dauter, Z., Dupont, C., Shareck, F., Morosoli, R., Kluepfel, D. & Davies, G. J. (2000). *J. Biol. Chem.* **275**, 23020–23026.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Esnouf, R. M. (1997). *J. Mol. Graph. Model.* **15**, 132–134.

Flint, H. J., Whitehead, T. R., Marin, J. C. & Gasparic, A. (1997). *Biochim. Biophys. Acta*, **1337**, 161–165.

Fontes, C. M. G. A., Gillbert, H. J., Hazlewood, G. P., Clarke, J. H., Prates, J. A. M., McKie, V. A., Nagy, T., Fernandes, T. H. & Ferreire, L. M. A. (2000). *Microbiology*, **146**, 1959–1967.

Fujimoto, Z., Kuno, A., Kaneko, S., Yoshida, S., Kobayashi, H., Kusakabe, I. & Mizuno, H. (2000). *J. Mol. Biol.* **300**, 575–585.

Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.

Harris, G. W., Jenkins, J. A., Connerton, I. & Pickersgill, R. W. (1996). *Acta Cryst.* D**52**, 393–401.

Harding, M. M. (2002). *Acta Cryst.* D**58**, 872–874.

Henrissat, B. & Davies, G. J. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.

Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. (1995). *FEBS Lett.* **362**, 281–285.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Kulkarni, N., Shendye, A. & Rao, M. (1999). *FEMS Microbiol. Rev.* **23**, 411–456.

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

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

Manikandan, K., Bhardwaj, A., Gupta, N., Lokanath, N. K., Ghosh, A., Reddy, V. S. & Ramakumar, S. (2006). *Protein Sci.* **15**, 1951–1960.

Miyazaki, K., Martin, J. C., Marinsek-Logar, R. & Flint, H. J. (1997). *Anaerobe*, **3**, 373–381.

Navaza, J. & Saludjian, P. (1997). *Methods Enzymol.* **276**, 581–594.

Otwinowski, Z. (1993). *Proceeding of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Parkkinen, T., Hakulinen, N., Tenkanen, M., Siika-aho, M. & Rouvinen, J. (2004). *Acta Cryst.* D**60**, 542–544.

Pell, G., Szabo, L., Charnock, S. J., Xie, H., Gloster, T. M., Davies, G. J. & Gilbert, H. J. (2004). *J. Biol. Chem.* **279**, 11777–11788.

Pell, G., Taylor, E. J., Gloster, H., Turkenburg, J. P., Fontes, C. M. G. A., Ferreire, L. M. A. Nagy, T., Clarke, J. H., Davis, G. J. & Gillbert, H. J. (2004). *J. Biol. Chem.* **279**, 9597–9605.

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.

Polizeli, M. L. T. M., Rizzatti, A. C. S., Monti, R., Terenzi, H. F., Jorge, J. A. & Amorim, D. S. (2005). *Appl. Microbiol. Biotechnol.* **67**, 577–591.

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

Roberge, M., Shareck, F., Morosoli, R., Kluepfel, D. & Dupont, C. (1997). *Biochemistry*, **36**, 7769–7775.

Schmidt, A., Schlacher, A., Steiner, W., Schwab, H. & Kratky, C. (1998). *Protein Sci.* **7**, 2081–2088.

Shallom, D. & Shoham, Y. (2003). *Curr. Opin. Microbiol.* **6**, 219–228.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.

Shulami, S., Zaide, G., Zolotnitsky, G., Langut, Y., Feld, G., Sonenshein, A. L. & Shoham, Y. (2007). *Appl. Environ. Microbiol.* **73**, 874–884.

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.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.

Usui, K., Suzuki, T., Akisaka, T. & Kawai, K. (2003). *J. Biosci. Bioeng.* **95**, 488–495.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *Protein Eng.* **8**, 127–134.

White, A. & Rose, D. R. (1997). *Curr. Opin. Struct. Biol.* **7**, 645–651.

White, A., Withers, S. G., Gilkes, N. R. & Rose, D. R. (1994). *Biochemistry*, **33**, 12546–12552.

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

Zechel, D. L. & Withers, S. G. (1999). *Acc. Chem. Res.* **33**, 11–18.

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