research papers

Structure determination of the extracellular xylanase from *Geobacillus stearothermophilus* by selenomethionyl MAD phasing

A. Teplitsky,[a] A. Mechaly,[b] V. Stojanoff,[c,d]* G. Sainz,[c] G. Golan,[a] H. Feinberg,[a] R. Gilboa,[a] V. Reiland,[a] G. Zolotnitsky,[b] D. Shallom,[b] A. Thompson,[e] 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, [b]Department of Biotechnology and Food Engineering and Institute of Catalysis Science and Technology, Technion – Israel Institute of Technology, Haifa 32000, Israel, [c]European Synchrotron Radiation Facility, ESRF, BP 220, Grenoble 38043, France, [d]National Synchrotron Light Source, BNL, Building 725D, Upton, NY 11973, USA, and [e]European Molecular Biology Laboratory, Grenoble Outstation, BP 181, Grenoble 38044, France

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

Xylanases are hemicellulases that hydrolyze the internal β-1,4-glycoside bonds of xylan. The extracellular thermostable endo-1,4-β-xylanase (EC 3.2.1.8; XT6) produced by the thermophilic bacterium *Geobacillus stearothermophilus* T-6 was shown to bleach pulp optimally at pH 9 and 338 K and was successfully used in a large-scale biobleaching mill trial. The xylanase gene was cloned and sequenced. The mature enzyme consists of 379 amino acids, with a calculated molecular weight of 43 808 Da and a pI of 9.0. Crystallographic studies of XT6 were performed in order to study the mechanism of catalysis and to provide a structural basis for the rational introduction of enhanced thermostability by site-specific mutagenesis. XT6 was crystallized in the primitive trigonal space group $P3_221$, with unit-cell parameters $a = b = 112.9$, $c = 122.7$ Å. A full diffraction data set for wild-type XT6 has been measured to 2.4 Å resolution on flash-frozen crystals using synchrotron radiation. A fully exchanged selenomethionyl XT6 derivative (containing eight Se atoms per XT6 molecule) was also prepared and crystallized in an isomorphous crystal form, providing full selenium MAD data at three wavelengths and enabling phase solution and structure determination. The structure of wild-type XT6 was refined at 2.4 Å resolution to a final R factor of 15.6% and an R_{free} of 18.6%. The structure demonstrates that XT6 is made up of an eightfold TIM-barrel containing a deep active-site groove, consistent with its 'endo' mode of action. The two essential catalytic carboxylic residues (Glu159 and Glu265) are located at the active site within 5.5 Å of each other, as expected for 'retaining' glycoside hydrolases. A unique subdomain was identified in the carboxy-terminal part of the enzyme and was suggested to have a role in xylan binding. The three-dimensional structure of XT6 is of great interest since it provides a favourable starting point for the rational improvement of its already high thermal and pH stabilities, which are required for a number of biotechnological and industrial applications.

1. Introduction

Glycoside hydrolases (GHs) hydrolyze glycoside bonds and are therefore implicated in a wide variety of biological processes. The potential use of these enzymes in the conversion of the large polymeric sugar biomass has fuelled the rapid growth of interest and activity observed recently in this field. Based on their sequence homologies, glycoside hydrolases are currently classified into more than 90 families (Henrissat & Davies, 1997, 2000), which are available on the constantly updated Carbohydrates Active Enzymes server (http://afmb.cnrs-mrs.fr/CAZY). Glycoside hydrolases use either a 'retaining' or an 'inverting' mechanism to hydrolyze the glycoside bond (Zechel & Withers, 1999) and can also be

sorted according to their mode of action, 'endo' or 'exo', which indicates whether the enzyme attacks the polysaccharide within the polymer chain or at one of its ends, respectively. Among the different GH families, distinctive variations have been observed in the active-site architecture and folds, most of which have been determined from three-dimensional structural studies (Davies & Henrissat, 1995).

Xylanases (β-1,4-D-xylan xylanhydrolase; EC 3.2.1.8) are glycoside hydrolases that attack the xylan backbone to produce both substituted and non-substituted shorter oligomers such as xylobiose and xylotriose. Xylanases have been isolated, characterized and classified mainly into two distinct glycoside hydrolase families, GH-10 and GH-11. GH-10 enzymes typically present an $(\alpha/\beta)_8$ fold (TIM-barrel) with a molecular weight of around 40 kDa, while GH-11 enzymes typically present a single domain of two highly twisted β-sheets in a 'jelly-roll' β-barrel configuration with a molecular weight of around 20 kDa (Davies & Henrissat, 1995).

Glycoside hydrolases, in general, and xylanases, in particular, have a wide range of uses and applications (Beg *et al.*, 2001; Subramaniyan & Prema, 2002; Shallom & Shoham, 2003). A major potential usage for xylanases is in the paper and pulp industry for biobleaching and biopulping (Viikari *et al.*, 1994; Tolan & Guenette, 1996), especially as environmentally friendly substitutes for chlorine-releasing processes. Because of the prevailing conditions in pulp processing, xylanases that are active at high temperatures and under high-pH conditions are of great industrial importance. A few years ago, we used the thermophilic bacterium *Geobacillus stearothermophilus* (formerly called *Bacillus stearothermophilus*) strain T-6 to isolate an extracellular xylanase that is naturally stable and active at relatively high temperatures and pH values. This xylanase was purified, characterized and cloned in *Escherichia coli* and was shown to bleach pulp optimally at pH 9 and 338 K (Khasin *et al.*, 1993; Gat *et al.*, 1994; Fishman *et al.*, 1995; Lapidot *et al.*, 1996; Mechaly, 1998; Shulami *et al.*, 1999). This enzyme was also successfully used in a large-scale biobleaching mill trial (Lundgren *et al.*, 1994; Rosenberg & Shoham, 1995), demonstrating its promising industrial potential.

The three-dimensional structure of the *G. stearothermophilus* T-6 xylanase (XT6) is of great interest because of its already high relative stability at elevated temperatures and under high-pH conditions. Moreover, this enzyme can be used as a good starting point for the design of a modified enzyme with even better industrial characteristics by systematic improvement of its stability and activity properties *via* rational protein engineering. In an attempt to solve the three-dimensional structure of XT6, we previously obtained single crystals of the native enzyme that are suitable for high-resolution X-ray diffraction studies. These crystals were also used to collect a full diffraction data set at room temperature (2.8 Å resolution) and at 103 K (2.3 Å resolution) (Teplitsky *et al.*, 1997). However, difficulties were encountered in our attempts to phase these native diffraction data and to determine the three-dimensional structure of the enzyme. We have recently had more success using the selenomethionyl MAD

phasing technique, which enabled us to determine the detailed crystal structure of wild-type XT6 at 2.4 Å resolution. In the current paper, we report the structure-determination procedure used for this analysis and describe the resulting three-dimensional structure of the enzyme. The biochemical significance of the structure is also discussed.

2. Materials and methods

Attempts to phase the XT6 crystallographic data using heavy-atom isomorphous replacement and molecular-replacement techniques were not successful. This failure could be explained, at least in part, by the lack of suitable binding sites for heavy atoms on the solvent-accessible surface of wild-type XT6 (for isomorphous replacement) and the lack of a suitable protein structure to be used as a reference model (for molecular replacement).

Two alternative experimental efforts have been undertaken in order to solve the crystallographic phase problem. In the first approach, a derivative of XT6 was prepared in which all eight methionine residues of the recombinant enzyme were substituted by selenomethionine residues (Se-XT6). As selenium gives a usable anomalous signal, such a derivative could be used for a multi-wavelength anomalous diffraction (MAD) phase determination (Hendrickson, 1985). Such a phase-determination technique has turned out to be very effective and successful in the last 10 years (Hendrickson & Ogata, 1997; Hendrickson, 1999; Ealick, 2000; Taylor, 2003; González, 2003; and references therein) and is predicted to become more commonly used in the future. In the second approach, a mutant of XT6 was prepared in which Glu265 in the active site was specifically replaced by a cysteine residue. This mutant was expected to provide a high-affinity binding site for heavy atoms (such as mercury and lead) in a solvent-accessible area of the enzyme. Hence it was expected to enable the production of a stable and specific heavy-atom derivative with high occupancy which would hopefully be sufficient for isomorphous replacement crystallographic phasing.

Although both experimental approaches turned out to be successful technically, it was the selenium derivatization that first resulted in interpretable maps. The procedures used for the Se-MAD structure determination are summarized in the following sections.

2.1. Production and characterization of Se-XT6

The preparation and characterization of the Se-XT6 derivative generally followed previously described procedures (Mechaly *et al.*, 2000). The extent to which selenomethionine had been incorporated into the protein was assessed by subjecting samples of the purified native enzyme and its selenomethionine derivative to electrospray mass-spectrometric analysis. The molecular-mass difference between selenomethionine XT6 and the native enzyme was determined to be 359.5. This value corresponds to an average content of 7.7 Se atoms per molecule of labelled protein and is in good agreement with the eight methionine residues present in XT6.

The observed data suggest a close to 100% substitution of methionine by selenomethionine residues.

To verify that the selenomethionine-containing enzyme adopts the structure of the native protein, we tested its catalytic activity. The enzymatic activities of XT6 and its selenomethionine derivative were determined using two different substrates: soluble oat spelt xylan and the chromogenic substrate para-nitrophenyl xylobiose (PNPX$_2$; Mechaly et al., 1997). The k_{cat} values for the selenomethionine XT6 by both assay methods were only about 35% less than that of the native enzyme. Also, both the native and the selenomethionine xylanases exhibited similar K_m values (1 mM) measured towards PNPX$_2$.

These kinetic data clearly demonstrate that the incorporation of selenomethionine into XT6 did not significantly affect its catalytic activity or its substrate specificity. The results support the observation that the majority of selenomethionyl proteins maintain their structure and catalytic activity (Hadener et al., 1993; Lee et al., 1996; Yang et al., 1990; Boles et al., 1991). In cases where the specific activity of the selenomethionine derivative was found to differ significantly from that of the native enzyme, a methionine residue was suggested to be involved either in catalysis (Graber et al., 1993) or in ligand binding (Frank et al., 1985).

2.2. Crystallization and preliminary X-ray analysis of Se-XT6

2.2.1. Crystallization. Recombinant native XT6 has previously been crystallized by the hanging-drop vapour-diffusion method (Teplitsky et al., 1997) and the crystals diffracted at room temperature to about 2.8 Å resolution (unit-cell parameters $a = b = 114.9$, $c = 122.6$ Å, $\alpha = \beta = 90$, $\gamma = 120°$; space group $P3_221$). The severe radiation damage observed at room temperature was partially solved by measurement of a full diffraction data set (to 2.3 Å resolution) in a cold nitrogen-gas stream (103 K) without a need for a cryoprotecting solution (unit-cell parameters $a = b = 112.2$, $c = 122.9$ Å, $\alpha = \beta = 90$, $\gamma = 120°$). Unfortunately, this type of flash-freezing experiment subsequently proved to be inconsistent and hence was not practical for a detailed structural analysis at high resolution.

Single crystals of the selenomethionine XT6 derivative were obtained using the hanging-drop method under conditions that differed slightly from those used for the native XT6 crystals. In general, a higher concentration of both protein and precipitant (ammonium sulfate) were necessary for crystallization and crystal growth was significantly slower than that of native XT6. For the best results, the initial drop content was 7–9 mg ml^{-1} protein, 0.9–1.0 M ammonium sulfate, 0.05 M sodium acetate buffer pH 5.0 and 4% dimethylsulfoxide (DMSO) as an organic additive. These drops were equilibrated against a 1 ml reservoir solution of 1.8–2.0 M ammonium sulfate.

The first crystals were observed after about two weeks and grew to their full size (typically elongated hexagonal prisms, $0.15 \times 0.15 \times 0.1$ mm) in four weeks. The fully grown crystals were usually smaller and exhibited a significantly higher sensitivity to air exposure and mechanical manipulations compared with the native XT6 crystals. Diffraction studies on these crystals at room temperature using X-ray synchrotron radiation (NSLS, Brookhaven National Laboratory, USA) indicated that the crystals of the selenomethionine XT6 are similar to the native crystals in unit-cell parameters ($a = b = 115.4$, $c = 124.3$ Å, $\alpha = \beta = 90$, $\gamma = 120°$), space group ($P3_221$) and diffraction properties (about 3.0 Å resolution limit).

2.2.2. Flash-cooling of the crystals. The high radiation-sensitivity of these crystals was overcome by flash-cooling techniques (at 103 K) using the original crystallization mother liquor containing 25%(w/v) glucose as a cryoprotecting solution. Crystals of Se-XT6 were flashed-cooled by the 'direct-freezing' procedure (see, for example, Zharkov et al., 2002; Gilboa et al., 2002; Golan et al., 2004). The crystals were taken from their original solution, immersed for about 30 min in 25%(w/v) glucose solution and then positioned quickly within a cold nitrogen-gas stream. This crystal-mounting procedure reproducibly resulted in radiation-insensitive crystals with isomorphous unit-cell parameters ($a = b = 113.39$, $c = 122.65$ Å) and the original $P3_221$ space group. Preliminary crystallographic analysis of the selenomethionine XT6 crystals (at both room temperature and 103 K) confirmed the existence of a significant selenium anomalous signal with a magnitude that should in principle be sufficient for selenium MAD crystallographic phasing. The preliminary diffraction data were measured on both an imaging-plate area detector and a CCD area detector using synchrotron radiation (X12C and X26C beamlines, NSLS, BNL).

2.3. Crystallographic structure determination

2.3.1. MAD data collection and processing. MAD crystallographic experiments for all XT6 crystals were performed on beamline BM14 in a procedure similar to that recently described for α-glucuronidase T-6 (GLT6; Golan et al., 2004). This beamline is installed on a bending magnet at the European Synchrotron Research Facility (ESRF, Grenoble, France). It uses a water-cooled silicon(111) monochromator, providing an energy resolution of less than 1 eV, which is appropriate for multi-wavelength anomalous diffraction (MAD) experiments. Data measurement for all crystals was performed in a nitrogen Cryostream (100 K, Oxford Cryosystems) using the crystal-freezing procedures described above [25%(w/v) glucose solution as a cryoprotectant]. A full MAD data collection around the selenium edge was performed with a single frozen crystal using a MAR345 imaging-plate area detector (MAR Research, Germany). Diffraction data were collected by the oscillation method at three carefully selected wavelengths: $\lambda_1 = 0.9791$ Å (the inflection point of the fluorescence spectrum, f'), $\lambda_2 = 0.9789$ Å (peak, f'' maximum) and $\lambda_3 = 0.9017$ Å (remote high-energy wavelength). After the completion of the MAD data collection at 2.9 Å resolution, a higher resolution (2.4 Å) data set has been collected on a native XT6 crystal in order to enable the full structure determination of the non-substituted

Table 1
Representative crystallographic data-collection parameters.

Values in parentheses are for the last shell.

Data set	Se-MAD† inflection	Se-MAD peak	Se-MAD remote	Native
Beamline	ESRF BM14	ESRF BM14	ESRF BM14	ESRF BM14
Wavelength (Å)	0.9791	0.9789	0.9017	1.0715
Space group	$P3_221$	$P3_221$	$P3_221$	$P3_221$
Unit-cell parameters				
$\quad a, b$ (Å)	113.39	113.39	113.39	112.87
$\quad c$ (Å)	122.65	122.65	122.65	122.73
$\quad \gamma$ (°)	120.0	120.0	120.0	120.0
Reflections				
$\quad$ Total	99311	99689	100670	161549
$\quad$ Unique	20597	20608	20709	35642
Redundancy	4.8	4.8	4.9	4.5
$\langle I \rangle / \langle \sigma(I) \rangle$	27.5	24.2	25.0	12.2
Mosaicity	0.72	0.71	0.74	0.42
Resolution range	25.0–2.90	25.0–2.90	25.0–2.90	25.0–2.40
(Å)	(2.95–2.90)	(2.95–2.90)	(2.95–2.90)	(2.45–2.40)
Completeness (%)	99.8 (99.7)	99.8 (99.9)	99.9 (100.0)	99.8 (100.0)
R_{sym} (%)	4.5 (12.6)	6.3 (14.8)	5.1 (14.9)	5.5 (21.1)

† Multi-wavelength anomalous diffraction (MAD) data collection at the selenium absorption edge.

enzyme. Data were processed using the programs *DENZO* and *SCALEPAC*K (Otwinowski, 1993; Otwinowski & Minor, 1997). The X-ray data-collection parameters and data-processing statistics for both the Se-XT6 and native XT6 are summarized in Table 1.

2.3.2. MAD phase calculation. The full data sets collected at three wavelengths around the selenium edge were used for MAD phasing, as well as for the initial refinement of the Se-XT6 structure. All further computation was carried out using the *CCP*4 program suite (Collaborative Computational

Project, Number 4, 1994). Data were scaled using the *ROTAPREP*, *AGROVATA*, *TRUNCATE* and *SCALEIT* programs. Anomalous difference and dispersive difference Patterson syntheses were calculated from the three MAD data sets combined together. These maps showed strong peaks at the expected Harker sections. The coordinates of the Se atoms were derived using the program *RSPS* (Knight, 2000). The *MLPHARE* routine (Otwinowski, 1991) was used for refinement of the initial estimates of the seven selenium positions and phase calculation. The space-group ambiguity ($P3_121$ *versus* $P3_221$) and the correct enantiomorph of the structure were uniquely resolved only by parallel phasing for the two alternatives in each case and corresponding examination of the agreement factors and the solvent-flattened final Fourier maps. This procedure unambiguously selected one of the possible enantiomorphs and indicated that the correct space group is $P3_221$. The best initial set of MAD phases with an overall figure of merit (FOM) of 0.57 were further improved by solvent flattening using the *CCP*4 program *DM* (Cowtan, 1994). The final electron-density map had an overall FOM of 0.89.

2.3.3. Model building. The initial electron-density map calculated at 2.9 Å resolution and using the MAD/*MLPHARE* phases was of relatively high quality (Fig. 1). Densities were clear (for 2.9 Å data) for both the main-chain and the side-chain regions of the protein. This allowed a straightforward main-chain tracing, side-chain interpretation and overall model building of the protein (performed mainly with *O*; Jones *et al.*, 1991). A linear representation of the electron density was initially calculated with the program *BONES* (Kleywegt & Jones, 1994). The α/β-barrel motif of the structure was immediately apparent in this linear-model representation of the protein.

The continuous linear model was then examined and transformed into a C^α protein skeleton based on the experimental electron density and standard backbone parameters. The resulting C^α skeleton was then examined with the BATON and LEGO utilities of *O* in order to build a poly-alanine model. Owing to the relatively high electron density of the Se atoms in the maps, it was possible to identify all eight selenomethionine sites at this stage. The correlation of these methionine sites together with the uniquely defined electron-density shape of several aromatic side chains (Tyr, Trp and Phe; Fig. 1) enabled us to trace and fit most of the protein residues (using *O*/LEGO) according to the known amino-acid sequence of XT6. Most amino acids (12–379) were sequentially incorporated into the emerging model of the protein, except for the first 11 amino acids at the amino-terminus,
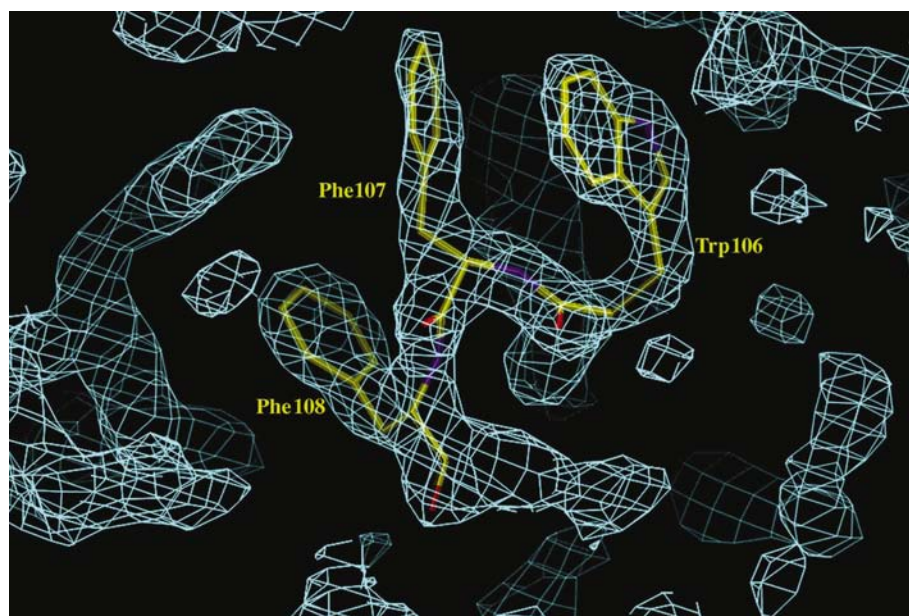

Figure 1
A representative section of the Se-XT6 experimental electron-density map (2.9 Å resolution, 1.0σ contour level, cyan) generated after the density-modification procedure. This section shows the density for Trp106, Phe107 and Phe108, which were also used for chain tracing and initial building of the model. Superimposed on the map is the corresponding region of the final model of Se-XT6 (stick representation, conventional atom colour codes).

which were difficult to fit owing to relatively poor electron density. Residues 5–11 were located at a later stage of the refinement, while residues 1–4 are still missing in the model, probably owing to relatively high flexibility and/or crystallographic disorder.

2.4. Structure refinement

2.4.1. Refinement of Se-XT6 at 2.9 Å resolution. Prior to the refinement of Se-XT6, 10% of the data were randomly flagged for cross-validation using R_{free} (Brünger, 1992; Kleywegt & Brünger, 1996). Crystallographic refinement was carried out against the 2.9 Å peak-wavelength data set, measured as part of the MAD experiment described above (Table 1). Simulated annealing was used with the program *CNS* (Brünger *et al.*, 1998), which converged to a crystallographic R factor of 29.9% and an R_{free} of 33.5%. After a few rounds of positional refinement using *X-PLOR* (Brünger, 1990) and *REFMAC* (Lamzin & Wilson, 1993), the entire molecule was inspected and adjusted using both $(2|F_o| - |F_c|)$ and $(|F_o| - |F_c|)$ maps. This procedure greatly improved the quality of the initial hand-built model, resulting in an improved Se-XT6 model with better connectivity and structural parameters.

2.4.2. Refinement of native XT6 at 2.4 Å resolution. The final model of the Se-XT6 derivative (at 2.9 Å resolution) was used for a full structural refinement against the 2.4 Å native data set, after changing all selenomethionine residues to methionines. From this stage on, all additional rounds of minimization and individual B-factor refinement were performed using the native XT6 structure, the native XT6 data

Table 2
Summary of XT6 structure refinement at 2.4 Å resolution.

Model refinement	
Data range (Å)	25.0–2.4
Data cutoff	$|F| > 0$
R factor† (%)	15.6
No. reflections	35080
R_{free}† (%)	18.6
No. reflections	3461
Refined model	
Residues	375
No. protein atoms	3071
No. solvent atoms	630
Average B (Å^2)	32.9
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.3
Solvent content (%)	76
No. molecules per AU	1
Estimated coordinate error (Å)	
Luzzati	0.20
σ_A	0.18

† $R = \sum ||F_o| - |F_c||/\sum |F_o|$. A random subset (9.9%) of the data was used for the calculation of R_{free}.

set and the *CNS* program (Brünger *et al.*, 1998). In the initial maps at this stage, it appeared as if the three peptide bonds between residues Glu209 and Pro210, between Trp242 and Pro243 and between Trp274 and Pro275 did not fit well into a *trans* geometry and they were suspected to be *cis*-peptide bonds. In the first few refinement rounds, we purposely treated those bonds as the standard *trans*-peptides in order to examine their true geometry, yet the refinement process forced these bonds to flip into the *cis* geometry, which indeed fitted better into the original electron density. Another round of refinement and calculation of a difference electron-density map confirmed the *cis* conformation of these three peptide bonds.

Ordered water molecules were assigned in a stepwise procedure (see, for example, Greenblatt *et al.*, 1997; Zharkov *et al.*, 2002; Golan *et al.*, 2004). Briefly, water molecules were added to the model after every cycle of refinement based on $(2|F_o| - |F_c|)$ difference maps contoured at the electron-density level of the standard deviation of the map (1σ contours). Similarly used were $(|F_o| - |F_c|)$ difference density maps contoured accordingly at 3σ levels. These water positions were further examined on the basis of their temperature (B) factors and hydrogen-bonding interactions with protein functional groups and/or other water molecules.

In later stages of the refinement and model adjustment, two regions of significant positive electron density were observed in the protein interior
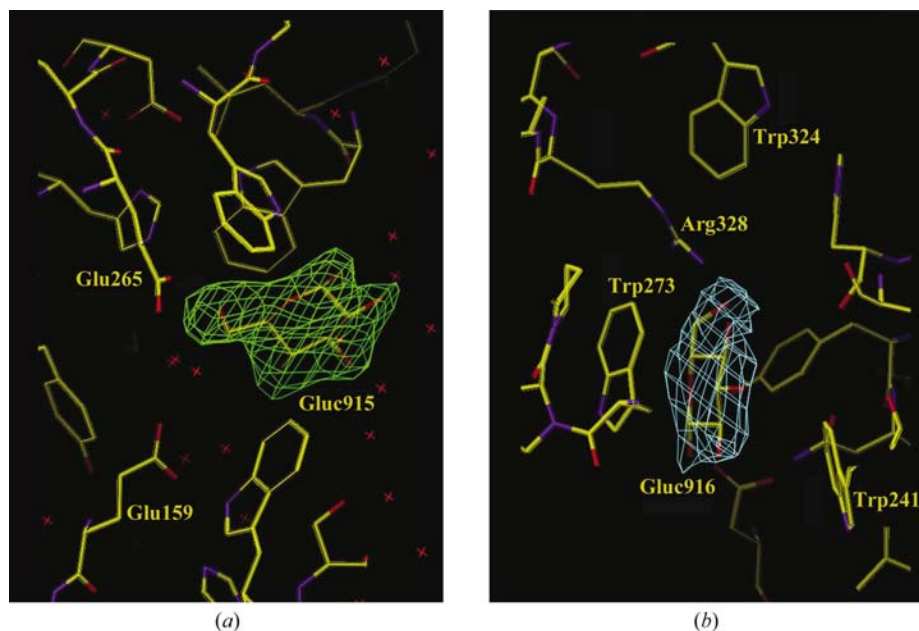

Figure 2
Omit map of the active site of native XT6 calculated after 20 cycles of refinement. The resulting electron density is shown here at a contour level of 3.0σ, demonstrating the unequivocal density for the two bound glucose molecules. Superimposed on the map is the corresponding region of the final model of native XT6 (conventional atom colors), including the glucose molecules. (*a*) The region of glucose 915. The two catalytic residues, Glu159 and Glu265, are indicated, as well as crystallographically identified water molecules (red crosses). (*b*) The region of glucose 916.

(Fig. 2). These regions, which are partially exposed to the solvent, were later identified as part of the binding site for oligoxylose substrates. The two defined difference electron densities in these regions could be perfectly fitted with two separated and independent glucose molecules, which were built into the model and refined with it during the final cycles of refinement. It is most likely that these glucose molecules diffused into the XT6 crystals during the 30 min soaking period in the cryoprotecting solution containing 25%(w/v) glucose (see below). The coordinates of a single glucose were obtained from the Hetero-compound Information Center Uppsala (HICup; http://alpha2.bmc.uu.se/hicup/). Parameter and topology files were set up using the existing carbohydrate-related files as templates. In the final refinement stages, it was also possible to identify a sulfate anion on the surface of the protein. Representative refinement details are summarized in Table 2.

The final refined model of native XT6 consists of 375 residues, beginning at residue 5 and ending at residue 379, the C-terminus of the protein. It also includes 630 water molecules, two glucose molecules and a single sulfate anion. Six amino-acid side chains (Glu42, Gln46, Glu137, Glu223, Ile228 and Asp308) were refined in two alternate conformations, with corresponding occupancies of 0.5 ± 0.05 for each conformation. This model gives an R factor of 15.6% and an R_{free} of 18.6% for all diffraction data to 2.4 Å resolution (no σ-cutoff). The overall quality of the final structure was examined using the program *PROCHECK* (Laskowski *et al.*, 1993). Analysis of the Ramachandran plot (Ramachandran *et al.*, 1963)
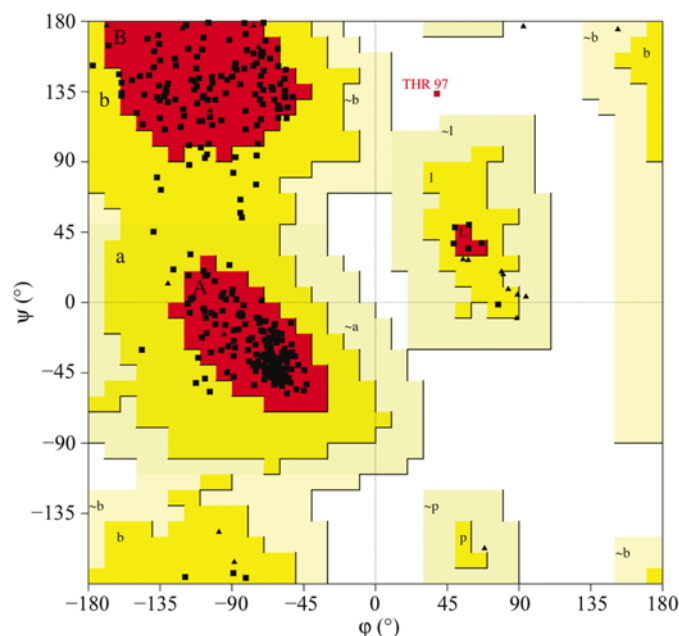
showed that 90.4% of the amino-acid residues are in the most favoured regions and 9.3% are in additionally allowed regions (Fig. 3). Only one residue (Thr97, 0.3%) is in a disallowed region. In general, bond lengths and bond angles conform closely to the standard values defined for protein structures by Engh & Huber (1991), with r.m.s. deviations of 0.006 Å and 1.3°, respectively. The average B value for the final structure of native XT6 is 32.9 Å^2 for all atoms, indicating relative rigidity of the molecule in the crystal and a reliable interpretation of the experimental electron-density map.

2.5. Calculations and figure preparations

The matrices for the superposition of the XT6/XlnA structures were calculated with a least-squares distance-minimization algorithm (*LSQ*, implemented within the program *O*) using the active-site C$^\alpha$ atoms as the guide co-ordinates. Figs. 1 and 2 were prepared using the programs *O* (Jones *et al.*, 1991) and *BOBSCRIPT* (Esnouf, 1997). Fig. 3 was prepared with the program *PROCHECK* (Laskowski *et al.*, 1993). Figs. 4(*a*), 4(*b*), 5(*a*) and 7 were prepared using a combination of the programs *MOLSCRIPT* (Kraulis, 1991), *BOBSCRIPT* (Esnouf, 1997) and *RASTER3D* (Merritt & Bacon, 1997). Fig. 5(*b*) was prepared using the program *LIGPLOT* (Wallace *et al.*, 1995). Fig. 6 was calculated and prepared using the program *CLUSTALW* (Thompson *et al.*, 1994) together with the ESPript server (Gouet *et al.*, 1999). Fig. 8 was prepared using the program *SPOCK* (Christopher, 1998).

3. Results and discussion

3.1. Structure determination of XT6

The crystallization and preliminary crystallographic characterization of native XT6 have been described in detail previously (Teplitsky *et al.*, 1997; Mechaly *et al.*, 2000) and summarized briefly above. Both isomorphous replacement and molecular-replacement attempts using the original diffraction data of native XT6 have failed in the structure determination efforts. Hence we switched to selenomethionine MAD phasing (of the Se-XT6 derivative) as the main approach towards solving the structure of this enzyme. The calculated MAD phases proved to be of sufficient quality to produce interpretable electron-density maps of Se-XT6 (at 2.9 Å resolution), which were further improved by density-modification (solvent-flattening) procedures. The resulting map pointed out the Se atoms and showed very clear features of the enzyme (Fig. 1), including secondary-structure elements and the solvent channels between them. After several rounds of refinement and model adjustment of the initial Se-XT6 structure at 2.9 Å resolution, a new model was obtained for native XT6. This model was then refined against the 2.4 Å diffraction data set to obtain the final structure reported and discussed here. The relatively high quality of the present structure is demonstrated in the clear identification of the bound glucose molecules (Fig. 2), the relatively good final crystallographic R factor and the normal distribution of side-

Figure 3
Ramachandran plot (Ramachandran *et al.*, 1963) for the final structure of XT6. Glycine residues are shown as triangles; all other residues are shown as squares. Most favoured regions (dark grey) are labelled with capital letters (A, B, L), additionally allowed regions (medium dark grey) are labelled in lower case (a, b, l, p) and generously allowed regions (light grey) are labelled with a tilde (~a, ~b, ~l, ~p). 90.4% of all residues are in the most favoured regions and 9.3% are in additionally allowed region. Only one residue (Thr97, 0.3% of the total) is in a disallowed region.

research papers

chain conformations in the Ramachandran plot (Fig. 3). Based on the resolution and *R* factor, the average experimental error in the coordinates of the final three-dimensional model of native XT6 should be around ±0.20 Å [according to the Luzzati error estimation (Luzzati, 1952) as implemented in the *CCP*4 software (Collaborative Computational Project, Number 4, 1994)], permitting a meaningful analysis of inter-actions and geometries in the structure presented here, even at this medium resolution (2.4 Å).

3.2. Unit-cell content

During the process of selenium-site localization, phase determination, model building and structure refinement, it was realised (to our surprise) that the crystallographic asymmetric unit of XT6 contains only one protein molecule. This corre-sponds to six molecules per unit cell rather than the 12 XT6 molecules that were expected on the basis of common Matthews coefficients (Matthews, 1968; Teplitsky *et al.*, 1997). This early conclusion of one monomer per asymmetric unit was supported by the identification of only seven selenium sites in the Patterson map (the Se atom of the first methionine at the N terminus of the molecule is usually difficult to locate) and the absence of non-crystallographic symmetry. In later stages of the refinement, the assignment of the asymmetric unit was unequivocally confirmed by the very clear electron-density distribution within the unit cell. The resulting solvent content of about 76% is indeed high and is not very common for protein crystals. Such a high water content (and the corresponding decrease in protein–protein contacts in the crystal) may explain, at least in part, the relatively high sensitivity of these crystals to X-ray radiation, cryosolvents and flash-freezing.

3.3. The overall structure

The overall structure of XT6 (Fig. 4) corresponds to a classical eightfold α/β-barrel with an additional amino-terminal α-helix and an additional atypical subdomain at the carboxy-terminus (Fig. 4*b*). The α/β-barrel (TIM-barrel) fold is a common architecture among enzyme structures (about 10% occurrence) and has been found in all family 10 xylanase struc-tures analyzed to date (*e.g.* Derewenda *et al.*, 1994; White *et al.*, 1994; Harris *et al.*, 1994, 1996; Dominguez *et al.*, 1995; Schmidt *et al.*, 1998; Natesh *et al.*, 1999; Lo Leggio *et al.*, 1999; Fujimoto *et al.*, 2000). As in most of the family 10 xylanases, the shape of the XT6 α/β-barrel is elliptical in cross-section (Fig. 4*a*), with the major axis running between β-strands 1 and 5 of the $(\alpha/\beta)_8$ domain. 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 which is about 15 × 11 Å wide and 18 Å deep.

A schematic topology diagram of the current structure of XT6 is shown in Fig. 4(*c*). The numbering of the α-helices (α*n*) and the β-strands (β*n*) was assigned in order to maintain the clas-sical arrangement of eight β-strands (β1–β8) in the internal core of the TIM-barrel $[(\alpha/\beta)_8$ fold]. The β-sheet strands of the TIM-barrel are accord-
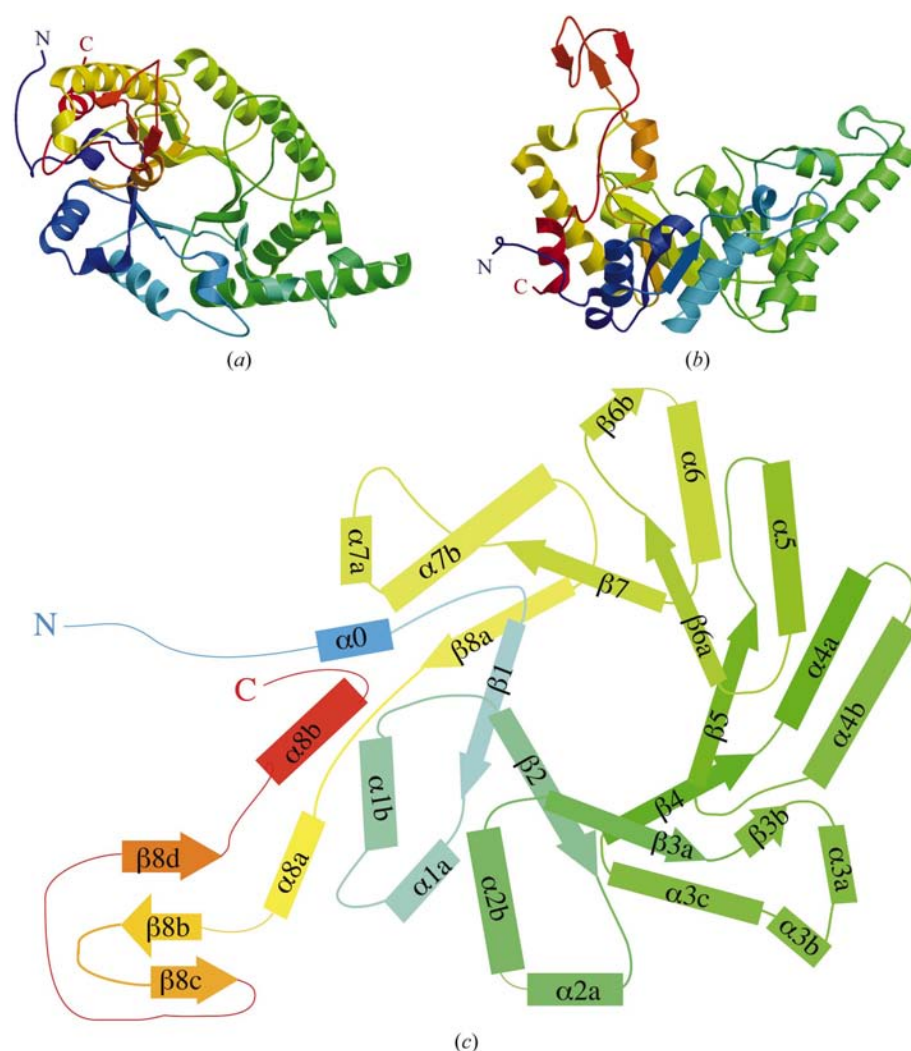

Figure 4
The overall structure of native XT6. The polypeptide chain of the enzyme is colour-coded so that it starts (amino-terminus, N) in dark blue and then gradually changes through blue, light blue, blue–green, light green, green, green–yellow, yellow, yellow–orange, orange and orange–red to red (carboxy-terminus, 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 family 10 xylanases. (*b*) A ribbon representation (side view) showing the unique subdomain (top left) at the C-terminal part of the enzyme. (*c*) A schematic topology diagram. α-Helices are presented as rectangles, β-strands as arrows and loops as curved lines. The numbers of the α-helices (α*n*) and β-strands (β*n*) are also indicated.

ingly constituted of residues 29–34 (β1), 54–57 (β2), 91–94 (β3a), 153–158 (β4), 197–202 (β5), 230–233 (β6a), 260–268 (β7) and 310–315 (β8a). Similarly, the α-helices of the TIM-barrel are constituted of residues 35–40 (α1a), 41–52 (α1b), 62–67 (α2a), 75–88 (α2b), 121–148 (α3c), 170–177 (α4a), 178–193 (α4b), 209–224 (α5), 243–256 (α6), 280–284 (α7a), 285–306 (α7b) and 369–377 (α8b). Still related to the α/β-barrel but outside the basic fold are the short β-strands constituted of residues 98–99 (β3b) and 236–238 (β6b) and the short α-helices constituted of residues 19–24 (α0), 104–108 (α3a) and 116–119 (α3b).

Unlike all the 'canonical' $(\alpha/\beta)_8$ folds found in other family 10 xylanases, the structure of XT6 contains an additional subdomain created by a short α-helix (α8a, residues 323–329) and three short β-strands (β8b, 331–333; β8c, 339–340; β8d, 349–351), all of which are connected by a relatively long sequence of loops (Fig. 4b). This atypical subdomain is located at the C-terminal end ('top') of the α/β-barrel and is inserted between the eighth β-strand (β8a) and the eighth α-helix (α8b). This subdomain spreads out of the molecule, away from the centre of the β-barrel and generally above the opening of the expected substrate-binding area (top left in Fig. 4b). The exact role of this subdomain is not yet known, but from its location in the vicinity of the active site it is likely to be involved in specific substrate binding. If true, this unique subdomain and these special substrate-binding interactions explain some of the unusual kinetics, substrate specificities and cleavage preferences found experimentally for XT6 (Zolotnitsky et al., in preparation).
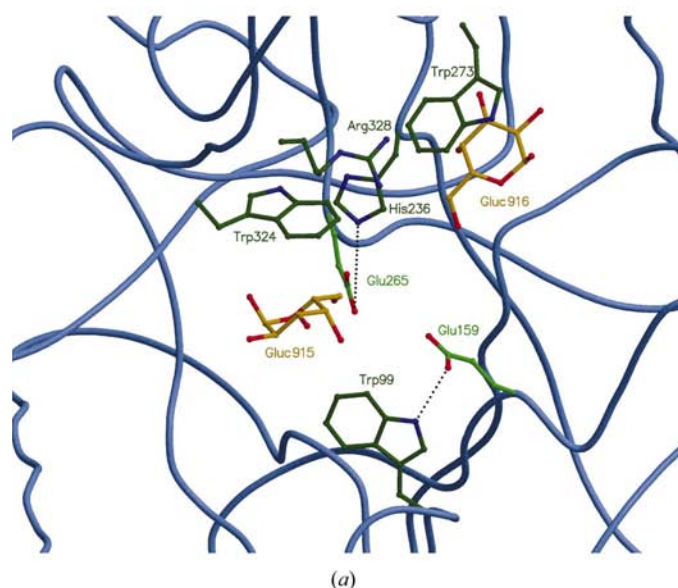
A long and open groove is clearly observed on the protein surface at the carboxy-terminal end of the β-barrel. This groove is exposed to the solvent and is lined by an array of aromatic and hydrophilic residues; its structure and dimensions correlate well with those of a xylan chain. As expected (and subsequently confirmed), this is the substrate-binding region, consistent with the location and the shape of the corresponding substrate-binding grooves found in homologous xylanases of family 10 (see below).

3.4. The active site

As in other xylanases, two glutamic acid residues (Glu159 and Glu265) are arranged on opposite sides of the surface groove, approximately 1/3 and 2/3 of the way along it, respectively (Fig. 5a). These acidic residues constitute the catalytic site of XT6 and the distance between their functional caboxylate groups (5.5 Å) is consistent with the expected 'retaining' hydrolytic mechanism (Derewenda et al., 1994;
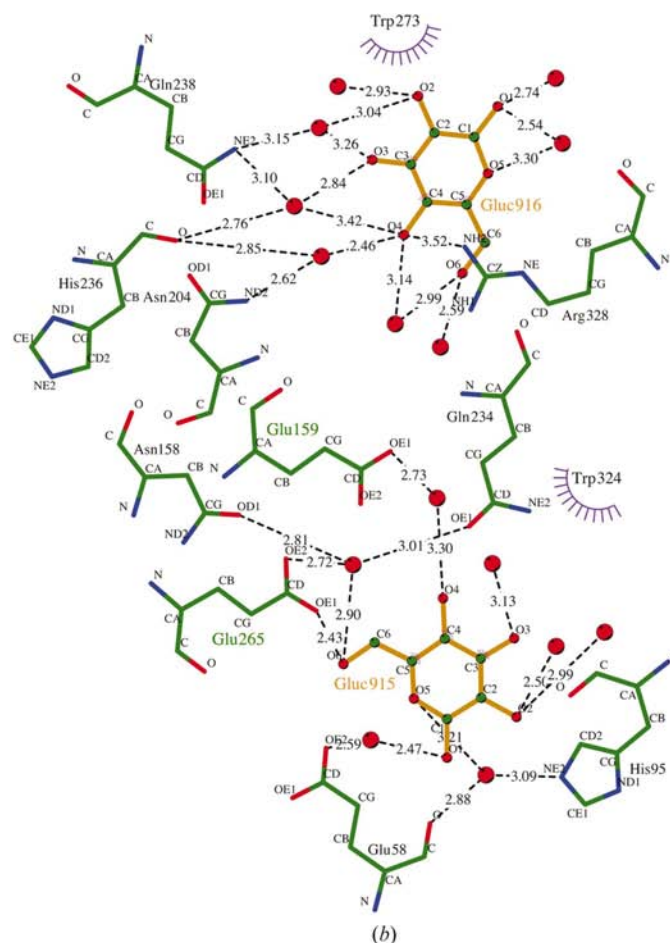

Figure 5
The active site of native XT6 as observed in the final refined model at 2.4 Å resolution. (a) A simplified representation of the active site in a top view, looking into the central cavity of the TIM-barrel. The general frame of the TIM-barrel is shown here as continuous coil (cyan), representing only the polypeptide main-chain atoms. The side chains of protein residues expected to be involved in substrate binding are shown in more details (stick model, usual atom colours). Similarly, the catalytic residues (C atoms in light green) and the two glucose molecules (Gluc915 and Gluc916; C atoms in yellow–green) are shown in detail. Dashed lines represent hydrogen bonds between the two catalytic residues and neighbouring side chains. (b) A schematic diagram of the interactions around the active site, including those of the two bound glucose molecules. The protein bonds are shown in green and the bound glucose bonds are shown in orange, while the rest of the atoms have the standard atom colours. Red spheres represent crystallographically determined water molecules. Dashed lines indicate hydrogen bonds or ionic interactions (distances are indicated in Å), while 'radiating' spheres indicate hydrophobic contacts between the bound glucose molecules (small spheres) and the neighbouring protein groups (larger spheres).

Davies & Henrissat, 1995; White & Rose, 1997; Zechel & Withers, 1999). Hydrolysis of the β-1,4-glycosidic bond by all known family 10 xylanases occurs *via* double displacement with two transition states and retention of configuration and is performed by two essential carboxylates located at the active site. One of these carboxylates usually functions as the catalytic nucleophile, while the other usually functions as the acid–base catalyst (White & Rose, 1997). Of these residues, the nucleophile usually has the most critical role, since it is directly involved in the formation of the glycosyl-enzyme covalent intermediate. In XT6, Glu265 is most probably the nucleophile, while Glu159 is most probably the acid–base catalyst, judged by their relative positions in the binding groove and the high correlation with both the sequences and structures of homologous family 10 xylanases (*e.g.* Derewenda *et al.*, 1994; Harris *et al.*, 1996; Schmidt *et al.*, 1998; Fujimoto *et al.*, 2000).

The carboxylate group of Glu159 is held in position by a hydrogen bond between its $O^{\varepsilon 1}$ atom and Trp99 $N^{\varepsilon 1}$ (2.9 Å) (Fig. 5*a*). The carboxylate group of Glu265 is held in position mainly by a direct interaction of its $O^{\varepsilon 1}$ atom with His236 $N^{\delta 2}$ (2.7 Å). The catalytic nucleophile (Glu265) is found on a β-bulge on the β7 strand, while the acid–base catalyst (Glu159) is found on a β-bulge at the C-terminal part of the β4 strand. Such a location of catalytic carboxylates is common for all glycoside hydrolases of the 4/7 superfamily, in which the two conserved glutamates are located near the carboxyterminal ends of β-strands 4 and 7 (Jenkins *et al.*, 1995).

Two glucose molecules, apparently derived from the glucose-containing cryoprotectant solution, were observed within the substrate-binding cleft, on opposite sides of the line connecting glutamates 265 and 159 (Fig. 5*a*). The experimental electron density is very well defined for both of these glucose molecules (Fig. 2). One of the glucose molecules (labelled 916) occupies a position approximately between the xylan-binding positions +1 and +2, using the conventional nomenclature for sugar-binding subsites in glycosyl hydrolases (Davies *et al.*, 1997). The other glucose molecule (labelled 915) similarly occupies a position approximately between the xylan-binding positions −1 and −2.

The presence and location of the glucose molecules in the active site of the present structure of XT6 could be explained by the high affinity of xylanases for sugars such as xylose and glucose. This sugar binding was probably also driven by the relatively long soaking (25–30 min) of the XT6 crystals in a cryoprotectant solution containing a relatively high concentration of glucose [25%(*w/v*)] prior to X-ray exposure. The two observed glucose molecules are surrounded by several aromatic residues, with possible stacking interactions (glucose 915 with Trp324 and glucose 916 with Trp273; Fig. 5*b*). Each of them is held in position by hydrogen bonds to amino-acid side chains of the active site. Glucose 915, found in the −1/−2 site, is hydrogen bonded to the catalytic nucleophile Glu265 and glucose 916, found in the +1/+2 site, forms two hydrogen bonds (although relatively weak) to Arg328 (Fig. 5*a*). The binding of the two glucose molecules observed in the active site of XT6 seem to mimic, to some extent, the binding of an oligosaccharide substrate. In this respect, the close proximity of Glu265 to O6 of glucose 915 may suggest that it makes catalytic interactions with the glycoside O atom of a real substrate. This proximity of the nucleophile to the target glycoside linkage constitutes the first step in the reaction mechanism proposed for retaining glycoside hydrolases (McCarter & Withers, 1994; Zechel & Withers, 1999).

The slightly different binding mode of the glucose molecules, both of which are not bound in 'proper' xylose-binding sites, could be a consequence of the different chemical structure (and hence different interactions) of glucose compared with xylose, the natural substrate of XT6. A slightly displaced xylose-binding site was observed for the xylanase from *Penicillium simplicissium* after a short soak in a solution containing monomeric xylose (Schmidt *et al.*, 1999), indicating that an oligomeric xylose (containing at least two sugar units) is needed for proper binding in the active site. These results suggest that the displacement of the two glucose-binding sites relative to the expected 'normal' xylose-binding sites in the current structure of XT6 could be also a result of the monomeric form of the glucose, in addition to its different chemical properties.

In a related structural analysis of XT6 mutant in which Glu265 has been changed to a cysteine residue (Teplitsky *et al.*, unpublished), only one bound glucose molecule (916) was observed in the active site. This molecule was found to have the same orientation and location (between the +1 and +2 sites) as the corresponding glucose molecule in the current structure of native XT6, but no electron density was observed for a glucose molecule in the −1/−2 binding site (two additional water molecules were instead located in this site). The absence of a second glucose molecule in the Glu265Cys mutant may be explained by the absence of a glutamic acid in position 265, the carboxylate group of which was shown to create a critical hydrogen bond with glucose 915 in the wild-type enzyme. This observation supports the suggested importance of the nucleophile (Glu265 in XT6) in the binding and positioning of substrates, reaction intermediates and their analogues.

3.5. Structural comparison of XT6 with other xylanases from family 10

It is interesting to compare the current structure of XT6 with related xylanases of family 10, especially in order to determine which structural features are conserved within the family and which features are unique to XT6. It is expected that most of the common features will indicate the general substrate binding and catalytic action within family 10 xylanases. Similarly, the unique structural features of XT6 are likely to include aspects related to substrate specificity and specificity-related mechanistic characteristics of XT6.

High-resolution crystal structure analyses of several family 10 xylanases have been reported in recent years. Such structures include, among others, those 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.*,

1994, 1996), the xylanase from *Clostridium thermocellum* (Dominguez *et al.*, 1995), the xylanase from *Pe. simplicissimum* (XlnA; Schmidt *et al.*, 1998, 1999) and the xylanase from *Thermosacus aurantiacus* (TAX; Natesh *et al.*, 1999; Lo Leggio *et al.*, 1999), all of which present the classical catalytic domain. Another structural analysis has been reported recently for the xylanase from *S. olivaceaviridis* E-86 (FXYN; Fujimoto *et al.*, 2000), which contains a xylan-binding domain in addition to the usual catalytic domain. A sequence alignment (Fig. 6) of selected xylanases of family 10 reveals a high

homology of amino acids within the family, especially around the active site and the inner parts of the catalytic core. The corresponding three-dimensional structures are therefore also expected to be very similar.

Indeed, all of the xylanases of family 10 which have been structurally analyzed to date have a catalytic core made up of an eightfold α/β-barrel (TIM-barrel). Only relatively small variations are observed between the various structures and these are usually located around the loops in the outer part of the α/β-barrel. The similarity of the common α/β barrel among
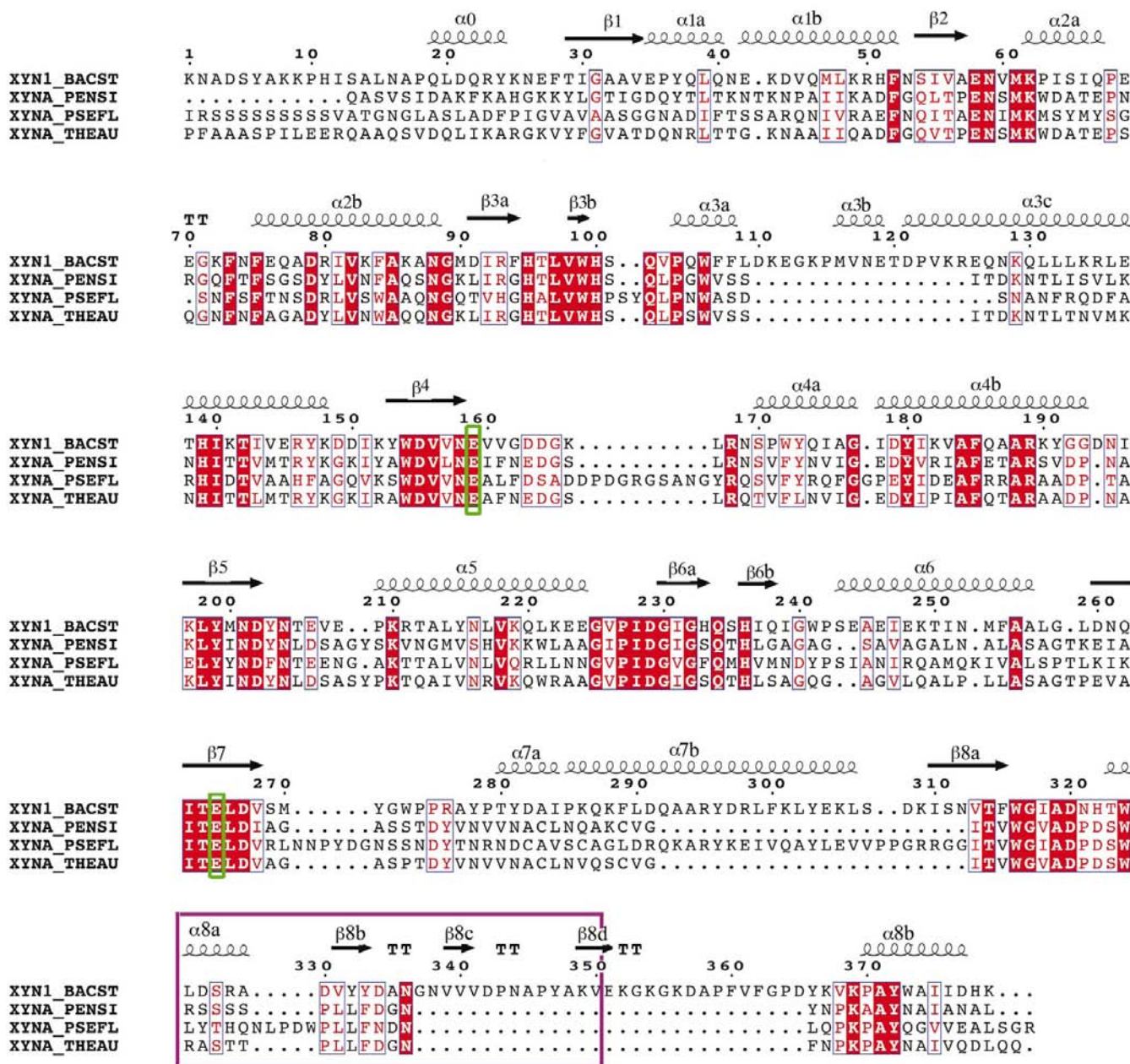

Figure 6
Amino-acid (single-letter codes) sequence alignment of family 10 xylanases computed with the program *CLUSTALW* (Thompson *et al.*, 1994). Included are the xylanase sequences of *Penicillium simplicissimum* (XYNA_PENSI), *Pseudomonas fluorescens* (XYNA_PSEFL) and *Thermoascus aurantiacus* (XYNA_THEAR), which present at least 30% sequence homology with XT6 (*G. stearothermophilus*, XYN1_BACST). The two catalytic glutamic acid residues (Glu159 and Glu265 in XT6) are marked in green frames. The full red squares represent full identity of the residues among the four species compared. The unique subdomain of XT6 is marked with a purple frame (residues 325–351). Secondary-structure elements of XT6 are marked above the relevant XT6 sequences (coils, α-helices; arrows, β-strands), correlated with the numbering scheme used in Fig. 4(*c*).

family 10 xylanases is demonstrated in Fig. 7(*a*), in which the current structure of XT6 is superimposed on the structure of the xylanase from *Pe. simplicissimum* (XlnA; Schmidt *et al.*, 1998). A common feature of all xylanases of family 10, also demonstrated in Figs. 7(*a*) and 7(*b*), is the small additional α-helix at the N-terminal of the molecule (α0 in XT6). As expected, the overall shape of family 10 xylanases is generally similar, yet local differences can be noticed in both the water-accessible surface of the molecule and the detailed charge distribution on it (Fig. 8).

While the shape and architecture of the active-site cleft is similar for both the xylanases from *G. stearothermophilus* (XT6) and the xylanase from *Pe. simplicissimum* (XlnA) (Fig. 7*c*), it is deeper and seems to be more negatively charged in the case of XT6 (Fig. 8). The higher rim along the elliptical long axis of the top cross-section observed for XT6 mostly

arises from the existence of the additional subdomain in the carboxy-terminal part of the enzyme, as mentioned above. In the overall structure of XT6 this subdomain is located close to the C-terminal end ('top') of the β-barrel, between the classical eighth β-strand and the eighth α-helix (Figs. 4*b*, 4*c* and 7*b*). This additional polypeptide chain starts with a short α-helix (α8a, seven amino acids), followed by a sequence of three short antiparallel β-strands (β8b, β8c and β8d), each of which contains only two or three amino acids. These elements extend from the main part of the α/β-barrel through a long bent loop. This distinct subdomain is a unique feature, so far found only for XT6 and certainly distinguishes it from all other family 10 xylanases analyzed to date. It is tempting to speculate about the flexibility of this domain, which is connected through elongated loops and could act as a 'lid' over the active-site cleft. The flexibility of this domain, which is also partially positively charged (Fig. 8), could allow a slightly differentiated mode of action of the active site, allowing XT6 not only to act as an 'endo' enzyme, but also to follow a processive action, as presented by certain cellobiohydrolases (Davies & Henrissat, 1995).

In addition to the novel subdomain discussed above, XT6 has several other small unique structural elements. These elements generally correspond to extra polypeptide segments (Fig. 6) which are not present in other xylanases of the family. Most of these extra features can be seen in Fig. 7, where XT6 is compared with XlnA (Schmidt *et al.*, 1998). The first such extra feature of XT6 is a 16 amino-acid segment (residues 110–125) that is not present in XlnA (Fig. 6). In both of these xylanases the third segment of the α-helices of the α/β barrel is divided into more than one part. In XlnA this segment starts with a short helix (corresponding to α3a in XT6) and then continues with a long helix (corresponding to α3c in XT6). The extra 16-residue segment of XT6 is inserted between these two parts of helix 3α, creating a long loop which separates α3a and α3c (Figs. 4*c* and 7*b*). This extra polypeptide chain in XT6 includes an additional short helix (α3b) and is also involved in elongating the α3c helix by five additional amino acids in the N-terminal part of the helix compared with XlnA. A similar phenomenon is observed when we compare the current structure of XT6 with the structures of the xylanases from *Ce. fimi* (White *et al.*, 1994),

Figure 7
Structural comparison of XT6 (green) with the xylanase from *Pe. simplicissimum* (XlnA, magenta). The superposition is based on the key atoms described in the text. (*a*) Top view showing the similarity of the two structures in the central (β/α)₈-barrel. (*b*) Side view showing the differences between the two structures, mainly around the unique subdomain of XT6 (top left). (*c*) A zoom into the active site of the two enzymes, demonstrating the almost identical arrangement of the two catalytic glutamic residues (Glu159 and Glu265 in XT6 and Glu132 and Glu238 in XlnA).
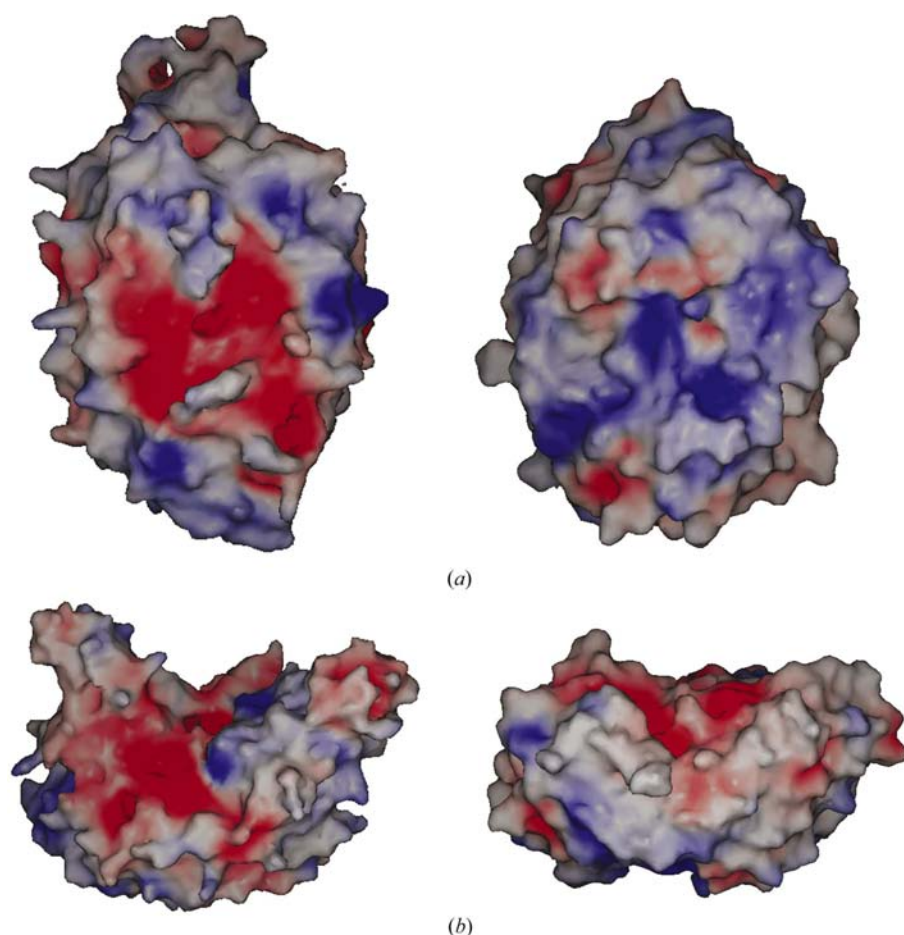
from *Ps. fluorescens* has a significantly longer connecting loop to the preceding $\beta 7$ strand, a loop that is stabilized by a calcium ion.

The exact functional meaning of the structural differences listed above for XT6, in comparison to other xylanases, is not completely clear at this point. A deeper understanding of these observations will definitely require additional experimental work, which will focus on the various structural elements discussed above. Especially interesting is the exact role of the unique subdomain, which should be investigated further by a combination of kinetic studies, mutagenesis and structural analysis utilizing various XT6 mutants and substrates. One such mutant of XT6, lacking the unique subdomain of the native enzyme, has recently been produced and is currently being studied in our laboratory.

Figure 8
The solvent-accessible surface of the xylanases from *G. stearothermophilus* (XT6, left) and *Pe. simplicissimum* (XlnA, right) coloured according to electrostatic potential (positive in blue, negative in red and neutral in grey). (*a*) Top view, demonstrating the differences in shape and charge distribution around the substrate-binding cavity. (*b*) Side view, showing the unique subdomain and the increased depth of the TIM-barrel in XT6.

Cl. thermocellum (Dominguez *et al.*, 1995) and *S. lividans* (Derewenda *et al.*, 1994). A comparison with the xylanase from *Ps. fluorescens* (Harris *et al.*, 1994, 1996) indicates that the inserted segment in this xylanase is shorter by about ten amino acids relative to that in XT6. As in XT6, this segment in the *Ps. fluorescens* xylanase creates a long loop between the $\alpha 3a$ and the $\alpha 3c$ subhelices, but the $\alpha 3c$ helix is significantly shorter (15 residues) in comparison to that in XT6 (28 residues). The longer $\alpha 3c$ helix is more typical of family 10 xylanases, at least those with a reported three-dimensional structure.

An additional difference observed in the current structure of XT6 compared with the xylanase from *Pe. simplicissimum* is the long $\alpha 7$ helix (built of two consecutive segments, $\alpha 7a$ and $\alpha 7b$), which is connected to the $\beta 7$ strand by a relatively long loop (Fig. 7*b*). Helixes corresponding to $\alpha 7$ of XT6 have similar length and orientation in the xylanases from *Ce. fimi*, *Cl. thermocellum* and *Ps. fluorescens*, compared with the $\alpha 7$ helixes in the xylanases from *Pe. simplicissimum* (Figs. 6, 7*a* and 7*b*) and *S. lividans*, which are significantly shorter. In this context, it is interesting to note that compared with all the other xylanases mentioned above, the $\alpha 7$ helix of the xylanase

References

Beg, Q. K., Kapoor, M., Mahajan, L. & Hoondal, G. S. (2001). *Appl. Microbiol. Biotechnol.* **56**, 326–338.
Boles, J. E., Cisneros, R. J., Weir, M. S., Odom, J. D., Villafranca, J. E. & Dunlap, R. B. (1991). *Biochemistry*, **30**, 11073–11080.
Brünger, A. T. (1990). *X-PLOR Version 3.0.* Yale University, New Haven, CT, USA.
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.

research papers

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

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.

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.

Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z., Wilson, K. S., Chauvaux, S., Beguin, P. & Alzari, P. M. (1995). *Nature Struct. Biol.* **2**, 569–576.

Ealick, S. E. (2000). *Curr. Opin. Chem. Biol.* **2**, 495–499.

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

Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–138.

Fishman, A., Berk, Z. & Shoham, Y. (1995). *Appl. Microbiol. Biotechnol.* **44**, 88–93.

Frank, P., Licht, A., Tullius, T. D., Hodgson, K. O. & Pecht, I. (1985). *J. Biol. Chem.* **260**, 5518–5525.

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

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

Gilboa, R., Zharkov, D. O., Golan, G., Fernandes, A. S., Gerchman, S. E., Matz, E., Kycia, J. H., Grollman, A. P. & Shoham, G. (2002). *J. Biol. Chem.* **277**, 19811–19816.

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.

González, A. (2003). *Acta Cryst.* D**59**, 1935–1942.

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

Graber, P., Bernard, A. R., Hassell, A. M., Milburn, M. V., Jordan, S. R., Proudfood, A. E. I., Fattah, D. & Wells, T. N. C. (1993). *Eur. J. Biochem.* **212**, 751–755.

Greenblatt, H. M., Almog, O., Maras, B., Spungin-Bialik, A., Barra, D., Blumberg, S. & Shoham, G. (1997). *J. Mol. Biol.* **265**, 620–636.

Hadener, A., Matzinger, P. K., Malashkevice, V. N., Louie, G. V., Wood, S. P., Oliver, P., Alefounder, P. R., Pittz, A. R., Abell, C. & Battersby, A. R. (1993). *Eur. J. Biochem.* **211**, 615–624.

Harris, G. W., Jenkins, J. A., Connerton, I., Cummings, N., Lo Leggio, L., Scott, M., Hazlewood, G. P., Laurie, J. I., Gilbert, H. J. & Pickersgill, R. W. (1994). *Structure,* **2**, 1107–1116.

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

Hendrickson, W. A. (1985). *Trans. Am. Crystallogr. Assoc.* **21**, 11–21.

Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.

Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.

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

Henrissat, B. & Davies, G. J. (2000). *Plant Physiol.* **124**, 1515–1519.

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.

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

Kleywegt, G. J. & Brünger, A. T. (1996). *Structure,* **4**, 897–904.

Kleywegt, G. J. & Jones, T. A. (1994). *Proceedings of the CCP4 Study Weekend. From First Map to Final Model,* edited by S. Bailey, R. Hubbard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.

Knight, S. D. (2000). *Acta Cryst.* D**56**, 42–47.

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

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

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

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 287–291.

Lee, Y.-H., Ogata, C. M., Pflugrath, J. W., Levitt, D. G., Sarma, R., Banaszak, L. J. & Pilkis, S. J. (1996). *Biochemistry,* **35**, 6010–6019.

Lo Leggio, L., Kalogiannis, S., Bhat, M. K. & Pickersgill, R. W. (1999). *Proteins,* **36**, 295–306.

Lundgren, K. R., Bergkvist, L., Hogman, S., Joves, H., Eriksson, G., Bartfai, T., van der Laan, J., Rosenberg, E. & Shoham, Y. (1994). *FEMS Microbiol. Rev.* **13**, 365–368.

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

McCarter, J. D. & Withers, S. G. (1994). *Curr. Opin. Struct. Biol.* **4**, 885–892.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Mechaly, A. (1998). PhD Thesis. Technion, Haifa, Israel.

Mechaly, A., Belakov, V., Shoham, Y. & Baasov, T. (1997). *Carbohydr. Res.* **304**, 111–115.

Mechaly, A., Teplitsky, A., Belakhov, V., Baasov, T., Shoham, G. & Shoham, Y. (2000). *J. Biotechnol.* **78**, 83–86.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

Natesh, R., Bhanumoorthy, P., Vithayathil, P. J., Sekar, K., Ramakumar, S. & Viswamitra, M. A. J. (1999). *J. Mol. Biol.* **288**, 999–1012.

Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering,* edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.

Otwinowski, Z. (1993). *Proceedings 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.

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

Rosenberg, E. & Shoham, Y. (1995). US Patent 5 434 071.

Schmidt, A., Gubiz, G. M. & Kratky, C. (1999). *Protein Sci.* **8**, 2403–2412.

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.

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

Subramaniyan, S. & Prema, P. (2002). *Crit. Rev. Biotechnol.* **22**, 33–64.

Taylor, G. (2003). *Acta Cryst.* D**59**, 1881–1890.

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.

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

Tolan, J. S. & Guenette, M. (1996). *Adv. Biochem. Eng. Biotech.* **57**, 289–310.

Viikari, L., Kantelinen, A., Sundquist, J. & Linko, M. (1994). *FEMS Micobiol. Rev.* **13**, 335–350.

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., Gilkers, N. R. & Rose, D. R. (1994). *Biochemistry,* **33**, 12546–12552.

Yang, W., Hendrickson, W. A., Kalman, E. T. & Crouch, R. J. (1990). *J. Biol. Chem.* **265**, 13553–13559.

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

Zharkov, D. O., Golan, G., Gilboa, R., Fernandes, A. S., Gerchman, S. E., Kycia, J. H., Rieger, R. A., Grollman, A. P. & Shoham, G. (2002). *EMBO J.* **21**, 789–800.