Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Characterization of a family 11 xylanase from *Bacillus subtillis* B230 used for paper bleaching

Aaron J. Oakley,[a] Tatjana Heinrich,[b] Colin A. Thompson[a] and Matthew C. J. Wilce[a]*

[a]School of Biomedical and Chemical Sciences, School of Medicine and Pharmacology, University of Western Australia, 35 Stirling Highway, Nedlands, WA 6009, Australia, and [b]Industrial Biosystems Pty Ltd, 14A Brennan Way, Belmont WA 6104, Australia

Correspondence e-mail: mwilce@receptor.pharm.uwa.edu.au

Received 12 December 2002
Accepted 14 January 2003

Enzymes such as family 11 xylanases are increasingly being used for industrial applications. Here, the cloning, structure determination and temperature-stability data of a family 11 xylanase, Xyn11X, from the alkali-tolerant *Bacillus subtilis* subspecies B230 are reported. This enzyme, which degrades xylan polymers, is being produced on an industrial scale for use in the paper-bleaching industry. Xyn11X adopts the canonical family 11 xylanase fold. It has a greater abundance of side chain to side chain hydrogen bonds compared with all other family 11 xylanase crystal structures. Means by which the thermostability of Xyn11X might be improved are suggested.

1. Introduction

Xylan is the major constituent of hemicellulose, a complex polysaccharide that is the most abundant biological carbon source in nature. Xylan is a polymer of 1,4-β-linked xylopyranosides, with various branching groups (depending on the source). Typically, the branching group is arabinose, glucuronic acid or acetate substituted at the 2- and 3-positions of the xylose moieties (Gruber *et al.*, 1998). Xylanases that degrade this polymer are found in bacteria, fungi and plants. The xylanases are part of a large family of glycosyl hydrolases, which have been classified into some 90 families based on their amino-acid sequences (Henrissat *et al.*, 1998). There are two main types of xylanase activities: (i) endo-xylanases (endo-1,4-β-xylanases; EC 3.2.1.8), which show a preference for internal xylan bonds, and (ii) exo-xylanases (exo-1,4-β-xylanases; EC 3.2.1.37), which show a preference for groups at the termini of xylan chains.

Of the glycosyl hydrolases families, the xylanases correspond to families 10 and 11. The family 10 xylanases have a molecular mass of approximately 35 kDa and possess the TIM-barrel $(\alpha/\beta)_8$ topology (Davies & Henrissat, 1995). Family 11 xylanases have a mass of approximately 22 kDa and a jelly-roll topology consisting of two twisted β-sheets stacked face to face. The structure has been likened to a right hand, with a two-β-strand 'thumb' forming a lid over the active site (Törrönen, 1994). The active site is thus located in the 'palm'. This topology has been found in plant lectins (Delbaere *et al.*, 1993) and in *Vibrio cholerae* neuraminidase (Crennell *et al.*, 1994). Of the family 11 xylanases, there are 11 representative structures from bacterial and fungal species (Table 1). Of these, the *Paecilomyces variotii*, *Thermomyces lanuginosus* and *Dictoglomus thermophilum* isozymes are thermostable, with the latter two having optimum activities at 343 and 348 K, respectively, and a catalytic pH optimum of 6.5 (Gruber *et al.*, 1998; McCarthy *et al.*, 2000).

Table 1
Xylanases with known tertiary structures.

Species (isozyme if applicable)	PDB code	Reference
Bacteria		
Dictyoglomus thermophilum	1f5j	McCarthy *et al.* (2000)
Thermomyces lanuginosus	1yna	Gruber *et al.* (1998)
Bacillus agaradhaerens	1qh6	Sabini *et al.* (1999)
Bacillus circulans	1xnb	Campbell *et al.* (1993)
Streptomyces sp. S38	1hix	Wouters *et al.* (2001)
Fungi		
Paecilomyces variotii	1pvx	Kumar *et al.* (2000)
Trichoderma harzianum	1xnd	Campbell *et al.* (1993)
Trichoderma reesei (xylanase II)	1enx	Törrönen *et al.* (1994)
Trichoderma reesei (xylanase I)	1xyn	Törrönen & Rouvinen (1995)
Aspergillus kawachii (xylanase C)	1bk1	Fushinobu *et al.* (1998)
Aspergillus niger (xylanase I)	1ukr	Krengel & Dijkstra (1996)

The use of xylanase in pulp and paper bleaching is of increasing interest (Kulkarni *et al.*, 1999). In the widely used kraft (or sulfate) process, cellulose fibres are separated from other wood components, such as xylan-containing lignin, by 'cooking' under strongly alkaline conditions. Approximately 10% of the lignin remains in the kraft pulp and are commonly removed by bleaching chemicals such as chlorine, which produce environmentally hazardous wastes (Neilson *et al.*, 1991). The use of xylanases holds the promise of reducing the amount of chlorine-based toxins in the pulp-bleaching process (Viikari *et al.*, 1994). It is believed that xylanase acts as a bleaching aid (bleach booster) by releasing trapped residual lignin within the pulp matrix, thus providing better access to bleaching chemicals. There has therefore been a great interest in engineering a highly stable xylanase for use in the paper-bleaching industry.

We report here the sequence, structure and thermostability data for a commercially produced xylanase, Xyn11X, which is a secretion product of *Bacillus subtilis* ssp. B230 isolated from white-rotted karri wood collected near Walpole, Western Australia. The enzyme is stable at temperatures of up to 343 K and shows optimal enzyme activity on pulp at 333 K and pH 8.0 (Dunlop *et al.*, 1996). The enzyme is being commercially marketed for use by the paper-pulp industry. Its structure has been determined to provide information for engineering to obtain improved temperature stability and to explore the structural and evolutionary basis of its thermostability.

2. Methods

2.1. Identification of the Xyn11X gene

The gene encoding B230 xylanase was identified as follows. The protein coded by Xyn11X was purified and the 25 N-terminal amino acids were deduced by Edman degradation. Degenerate PCR primers were designed using the deduced N-terminal DNA sequence: xynfor2 (5′-ATT ACB TCW AAT CAR ACK GGB AC) and the EGYQSSG motif common to all family G xylanases, xynrev3 (5′-CCR CTR CTT TGR TAS CCY TC). (The definitions of the degenerate nucleotides are as follows: B = G or T or C, K = G or T, R = A or G, S = G or C, W = A or T and Y = C or T.) PCR of B230 genomic DNA with these primers resulted in a 558 bp DNA fragment. The fragment was cloned into pGemT (Promega) and the DNA sequence determined (PRISM dye terminator cycle sequencing kit with FS polymerase; Perkin Elmer). The start of the gene sequence (named Xyn11X) matched the deduced DNA sequence from protein sequencing. The deduced amino-acid sequence from the internal Xyn11X fragment showed 91% identity with an internal fragment from *B. pumilis* XynA (Fukusaki *et al.*, 1984).

Flanking DNA sequences were isolated by inverse PCR. For isolation of the upstream region, B230 genomic DNA was digested with *Hin*dIII, self-ligated at 289 K and subjected to PCR with inverse primers xfor2 (5′-GGA ACC GTA TCT GTC AGC GAA C) and xrev1 (5′-CCA CTG GGC ACT GAA CGC ACC). A 950 bp PCR fragment was purified, cloned into pGemT and sequenced. The sequence contained a 438 bp ORF which showed high homology to the carboxy-terminus of the *B. subtilis xylB* gene, an intergenic region with several potential promoter motifs, a typical *B. subtilis* ribosome-binding site and the coding region for a 27 amino-acid leader peptide.

The downstream region was isolated by inverse PCR on *Eco*RI-digested self-ligated B230 genomic DNA with PCR primers xfor3 (5′-CCG CGC TTA CGG TAG AAG GC) and xrev2 (5′-CGT TCC CGT CGG TCT GTA GGT GC). The PCR products were cloned and sequencing revealed a known part of the Xyn11X coding region, 114 bp coding for the last 38 amino acids of Xyn11X, plus an intergenic region of 190 bp and the start of an ORF which showed 85% identity with the *B. subtilis pps* (phosphoenolpyruvate synthase) gene.

To confirm the data obtained from these three cloning steps, primers to the end of the *xylB* gene (endXylB; 5′-AAT GAT AAG TGT GGT TCG TAA TGG) and the start of the *pps* gene (PEP-rev1; 5′-TCC AGT TCA TGT AAA CCA AGT ACC) were used for PCR on B230 genomic DNA. A PCR fragment of the expected size of 1277 bp was obtained and cloned into pGemT. Sequencing of three independent clones in both directions with six different sequencing primers confirmed earlier results, except for one change at the end of the *Xyn11X* gene, which resulted in one change in the deduced amino-acid sequence. The 696 bp sequence encoding Xyn11X and its amino-acid translation is shown in Fig. 1. The sequence features a leader peptide of 27 amino acids that is absent in the mature form of the protein.

2.2. Protein purification

B. subtilis B230 was grown in LB medium containing 0.5%(*w*/*v*) xylose for 22 h at 310 K. The bacteria were pelleted and the supernatant was subjected to ammonium sulfate precipitation. The precipitate at 40%(*w*/*v*) ammonium sulfate was pelleted by centrifugation at 19 000*g* for 20 min. The pellet was resuspended in the initial volume of water and dialysed three times against water to remove ammonium sulfate. The solution was vacuum-dried and the protein

content of the powder was determined to be 28% (Bradford protein assay, BioRad).

Purification of B230 xylanase was completed by preparative polyacrylamide gel electrophoresis using a native 6% gel with a running solution of 25 m*M* histidine, 30 m*M* 3-*N*-morpholinopropanesulfonic acid buffer pH 6.5. B230 xylanase has a pI of approximately 8.8 and is thus positively charged at pH 6.5. Therefore, the electrophoresis was run with reversed poles. The first major peak after the elution front contained the xylanase. SDS–PAGE with silver staining revealed no other proteins in these fractions. The protein content of the two peak fractions was between 7 and 16 mg ml^{-1} over two runs. The resulting protein solutions were pooled and used in crystallization experiments.

2.3. Crystallization

Purified protein (10 mg ml^{-1} in 25 m*M* histidine, 30 m*M* 3-*N*-morpholinopropanesulfonic acid buffer pH 6.5) was subjected to crystallization trials using the sparse-matrix screen Hampton Crystal Screen I (Hampton Research, CA, USA) in the hanging-drop configuration. Thin plates were found to grow in condition 15: 0.2 *M* ammonium sulfate, 0.1 *M* cacodylate pH 6.5 and 30% polyethylene glycol (PEG) 8000. Improved crystals were obtained upon screening different pH and PEG concentrations. The final conditions used in this study were 20–40% PEG 8000, 0.2 *M* ammonium sulfate and 0.1 *M* cacodylate pH 6.3–6.7. Crystals grew as thin rectangular plates up to 0.1 mm in length within one week. The hanging drop was formed by the addition of 2 µl protein solution to 2 µl of well solution in all experiments.

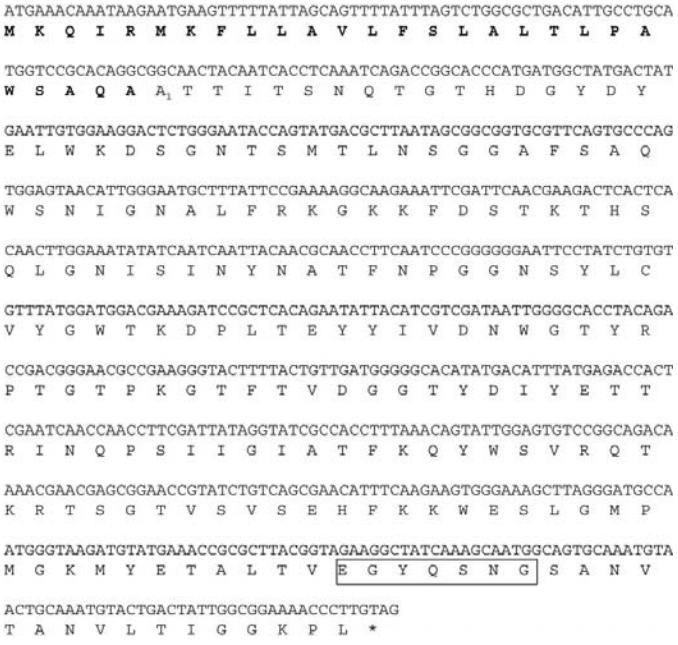

Figure 1
Nucleotide sequence of the Xyn11X-coding region (696 bp) and amino-acid translation (232 amino acids). The leader peptide of 27 amino acids is shown in bold. Numbering of the protein starts with Ala1 of the mature protein. The sequence equivalent to the family 11 'EGYQSSG' motif is boxed.

Table 2
Crystallographic data.

Statistics over all data are shown, with corresponding data in the highest resolution bin 2.28–2.20 Å shown in parentheses.

Space group	$P2_1$
Unit-cell parameters (Å, °)	a = 49.19, b = 52.54, c = 65.41, β = 97.83
No. of observed reflections	36059 (3107)
No. of unique reflections	16265 (1479)
Redundancy	2.22 (2.10)
R_{sym} (%)	10.4 (30.9)
$I/\sigma(I)$	7.67 (2.38)
Completeness (%)	95.0 (87.1)
Completeness > 3σ (%)	67.2 (20.9)

2.4. X-ray data collection

Crystals were snap-frozen in their native mother liquor using an Oxford cryocooler operating at 100 K. Data were collected using a MAR Research area detector mounted with focusing mirrors on a RU-200 rotating-anode X-ray generator producing Cu $K\alpha$ (1.54 Å) radiation. Reflections corresponding to Bragg spacings as small as 2.2 Å were observed. The data were indexed and processed using the *HKL* package (Otwinowski & Minor, 1997). The crystallographic statistics are indicated in Table 2. Intensities were converted to structure factors using *TRUNCATE* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994).

2.5. Structure solution and refinement

Structure-factor phases were determined using the molecular-replacement method as implemented in *MOLREP* (Vagin & Teplyakov, 2000). The structure of *B. agaradhaerens* xylanase (Sabini *et al.*, 1999; PDB code 1qh6) was used as the search model. (The sequence identity of *B. agaradhaerens* and B230 xylanase protein sequences is 73%.) *O* (Jones *et al.*, 1991) was used to visualize electron-density maps and rebuild the model. *CNS* (Brünger *et al.*, 1998) was used for all model refinement and electron-density map calculations.

2.6. Structure analysis

Coordinates for related xylanases were obtained from the Protein Data Bank (http://www.rcsb.org). *LSQMAN* (Kleywegt, 1996) was employed to structurally superimpose these xylanases. A structure-based sequence alignment of the aligned xylanases was then made using *STAMP* (Russell & Barton, 1992) and residues with a probability of residue structural equivalence (P_{ij}) greater than 0.5 were identified. Intra-molecular hydrogen bonding was analysed using *HBPLUS* (McDonald & Thornton, 1994). Surface-area composition by residue was calculated using *GRASP* (Nicholls *et al.*, 1991). Clustering of atoms within the xylanase structures was analysed using *ClusterOne* (A. J. Oakley, unpublished program).

2.7. CD spectra data collection

0.5 ml of xylanase purified as above was dialysed using a 3500 Da molecular-weight cutoff Slide-A-Lyzer (Pierce Chemical Co.) overnight against 1000 ml of 10 m*M* K_2HPO_4 buffer pH 8.0 to remove any non-proteinaceous UV-active compounds. The protein was diluted to a final concentration of 0.25 mg ml^{-1}. The sample was placed in a 2 mm quartz cell and subjected to circular dichroism (CD) spectropolarimetry using a Jasco J-810 spectrophotometer equipped with a Peltier temperature-control unit. CD spectra between 190 and 250 nm were measured at 10 K intervals from 293 to 323 K inclusive and then at 1 K intervals starting at 328 K and ending at 351 K. CD at 217 nm (corresponding to a trough in the spectrum) was monitored at every 1 K increment between 293 and 353 K. CD spectra were analysed using *CDPro* (Sreerama & Woody, 2000).

2.8. Phylogenetic analysis

Sequences of family 11 xylanases were obtained from the GenBank database (http://www.ncbi.nlm.nih.gov). Incomplete, truncated and redundant sequences were removed, leaving 97 family 11 xylanase sequences. A phylogenetic tree was produced using the bootstrapped neighbour-joining method as implemented in *CLUSTAL* (Thompson *et al.*, 1994).

3. Results

The diffraction data extend to a resolution of 2.2 Å (Table 2). The asymmetric unit contained two monomers (Matthews coefficient = 1.86 $Å^3$ Da^{-1}). Using *B. agaradhaerens* xylanase as the search model, the first solution found corresponded to the highest peak in the rotation and translation functions. The first solution was fixed and a second solution, corresponding to the second highest peak of the rotation function and the highest peak from a second translation search, was found. The molecular-replacement statistics are shown in Table 3.

The starting model had an *R* factor of 41.6% (R_{free} = 43.4%). This fell to 41.0% (R_{free} = 43.0%) after rigid-body refinement in *CNS* (Brünger *et al.*, 1998). Each monomer was treated as a rigid body. After positional refinement, the *R* factor was 31.5% (R_{free} = 35.0%). The model was subjected to five cycles of rebuilding in *O* (Jones *et al.*, 1991) (during which the residues in the model were changed from those in the *B. agaradhaerens* xylanase sequence to the *B. subtilis* B230 xylanase sequence), using σ_A-weighted $2F_o - F_c$ and $F_o - F_c$ maps. After each rebuild, the model was subjected to a round of positional and temperature-factor refinement in *CNS*. At the fourth rebuild, water molecules and a sulfate group were built into the model. Further water molecules were built and after the fifth cycle 109 waters and one sulfate group were present. The final model has an *R* factor of 20.8% (R_{free} = 27.2%). Throughout the refinement, non-crystallographic symmetry (NCS) restraints were employed and electron density was checked to determine whether any NCS breakdown had occurred. Where differences in conformation of NCS-related residues were observed, NCS restraints were released. The statistics for the final model (Fig. 2) are shown in Table 4.

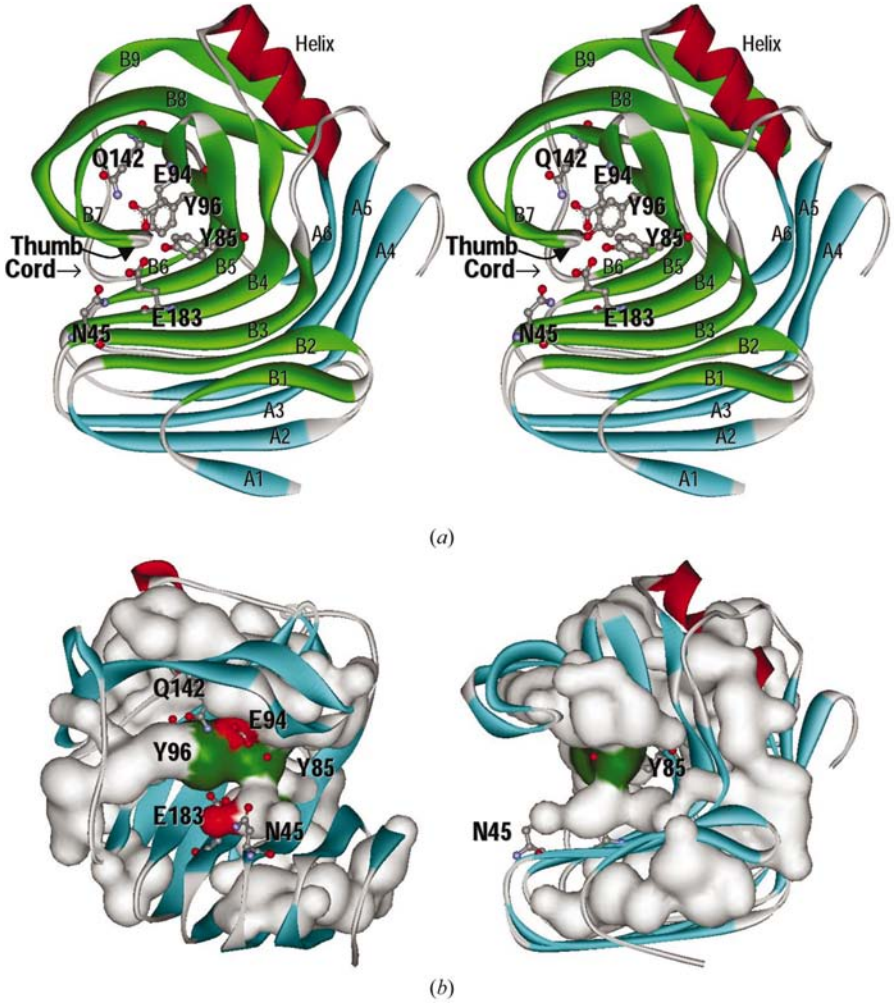

Figure 2
(*a*) Ribbon representation of Xyn11X. Each element of secondary structure is labelled. Sheet *A* is represented in cyan and sheet *B* is represented in green. Active-site residues (labelled) are shown in ball-and-stick representation. The 'cord' and 'thumb' motifs found in all family 11 xylanases are indicated. (*b*) Orthogonal views of ribbon diagram of Xyn11X with atoms forming hydrophobic clusters (as designated by *ClusterOne*) represented as a surface. All atoms are coloured white except for the catalytic residues (red) and the supporting tyrosine residues (green).

Table 3
Molecular-replacement statistics from *MOLREP*.

(*a*) Rotation-function solution.

	α, β, γ (°)	Peak height (σ)
Solution 1	190.01, 81.34, 22.61	13.41
Solution 2	11.97, 65.01, 18.02	12.80

(*b*) Translation-function solution.

	T_x, T_y, T_z (unit-cell fractions)	*R* factor	Correlation
Solution 1	0.346, 0, 0.395	0.512	0.359
Solution 2 (with solution 1 fixed)	0.660, 0.981, 0.828	0.439	0.550

Table 4
Refinement statistics.

Values in parentheses refer to the highest resolution bin (2.34–2.2 Å).

No. of atoms	
Protein	3186
Water	109
Sulfate	5
R factor	20.8 (27.3)
R_{free}	27.2 (33.9)
Resolution range (Å)	20–2.2
Reflections (overall)	16199 (2593)
Completeness (%)	95.4 (92.1)
Reflections in test set	832 (129)
Completeness (%)	5.0
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.006
Bond angles (°)	1.4
Dihedral angles (°)	26.4
Improper angles (°)	0.69
Rm.s.d. on bonded *B* factors (Å^2)	3.88
Residues in most favourable region of Ramachandran plot (%)	88.4

A structure-based sequence alignment of family 11 xylanases is shown in Fig. 3, with comparison statistics given in Table 5. Overall, most structural features of the xylanases, including the active site, are highly conserved. The highest r.m.s. deviation of any pair of xylanase structures is 1.9 Å (over 154 matching C^{α} atoms) between *B. circulans* and *B. agaradhaerens* xylanases. Table 6 shows the residues contributing to salt bridges in thermostable xylanases. *T. lanuginosus* xylanase appears to have the most salt bridges. The results of *HBPLUS* analysis of hydrogen bonds for all family 11 xylanase crystal structures are shown in Table 7. Of the xylanases analysed, Xyn11X had the most hydrogen bonds, *T. lanuginosus* xylanase had the greatest percentage of accessible surface area composed of charged residues and *A. kawachii* and *A. niger* had the highest percentage of accessible surface residues composed of polar residues.

Phylogenetic analysis of the sequences of known xylanase sequence reveals several distinct clades (Fig. 4). Multiple clades possess xylanases that are known to be thermostable. The closest relatives of Xyn11X appear to be *Bacillus* sp. YA-335 XYNY and *B. pumilus* HB030 XYNA.

The CD spectra of xylanase measured at 293 K is shown in Fig. 5(*a*). The measurement of CD at 217 nm *versus* temperature is shown in Fig. 5(*b*). The loss of CD signal indicates that the protein molecules start to denature at around 335 K with an apparent T_m of 340 K. Analysis with *CDPro* predicts that xylanase is 49% β-strand and 1.5% α-helix. This contrasts with *DSSP* (Kabsch & Sander, 1983) analysis of the coordinates (62% sheet and 6.8% helix). The

Figure 3
Structure-based sequence alignment of representative family 11 xylanases. T. reesei, *Trichoderma reesei* xylanase I (PDB code 1ynx); A. kawachii, xylanase C from *Aspergillus kawachii* (1bk1); A. niger, *Aspergillus niger* xylanase I (1ukr); B. circulans, *Bacillus circulans* xylanase (1xnb); T. lanuginosus, *Thermomyces lanuginosus* xylanase (1yna); P. variotii, *Paecilomyces variotii* xylanase (1pvx); T. harzianum, *Trichoderma harzianum* xylanase (1xnd); T. reesei II, *Trichoderma reesei* xylanase II (1enx); D. thermophilum, *Dictyoglomus thermophilum* xylanase (1f5j); B. subtilis B230, the subject xylanase of the current study; B. agaradhaerens, *Bacillus agaradhaerens* xylanase (1qh6); S. sp. S38, *Streptomyces* sp. S38 (1hix). Residues that show structural similarity ($P_{ij} > 0.5$) are capitalized. The catalytic acidic residues are indicated with an asterisk), with the catalytic tyrosine residues indicated (O). Other crucial active-site residues are also indicated (^). The secondary structure as determined by *DSSP* for Xyn11X is shown above the sequences. Sheet *A* is coloured cyan. Sheet *B* is coloured green. The α-helix is indicated in red. The amino acids A, F, M, I, L, V, W and Y are coloured yellow, Q, N, S and T are coloured green, E and D are coloured red, H, K, and R are blue and P, G and C are white.

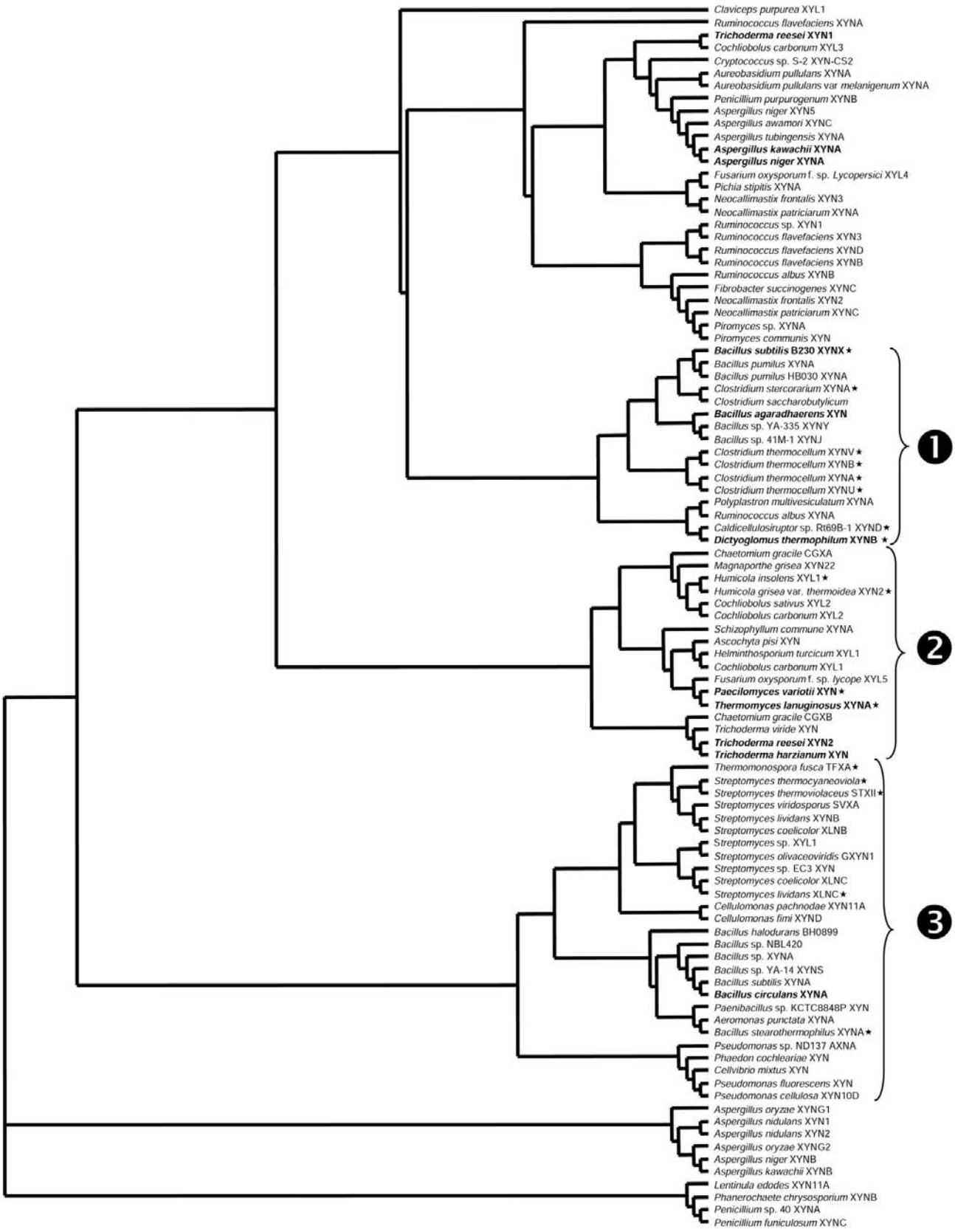

Figure 4
Phylogenetic tree calculated using the neighbour-joining method. Xylanses for which three-dimensional structures have been determined are shown in bold. Three clades that are known to contain thermostable or thermotolerant xylanases are indicated. Thermostable xylanases (or xylanases from thermostable organisms) are indicated with a star.

Table 5
Structure comparison data.

Structural identities expressed as r.m.s. values (Å) of structurally aligned xylanases. Xylanases were aligned in *LSQMAN* with a distance cutoff of 3.8 Å for determining structural similarity. The number of residues in each alignment are shown beneath the r.m.s. values. The total number of residues in each structure is given in parentheses. The PDB files used are denoted in the legend to Fig. 3.

	12	11	10	9	8	7	6	5	4	3	2	1
T. reesei I	1.373	1.298	1.260	1.398	1.243	1.327	1.299	1.400	1.271	1.037	1.069	
	169	168	165	167	170	169	173	174	171	168	168	(178)
A. kawachii	1.122	1.332	1.158	1.386	1.236	1.185	1.284	1.593	1.334	0.228		
	161	164	159	165	170	166	168	171	164	181	(183)	
A. niger	1.209	1.310	1.298	1.553	1.165	1.197	1.228	1.438	1.287			
	164	162	163	167	167	166	170	167	163	(183)		
B. circulans	0.962	1.952	1.038	1.271	1.286	0.907	1.185	1.327				
	169	154	168	165	164	174	174	174	(185)			
T. lanuginosus	0.972	1.132	1.219	1.126	0.781	0.964	0.384					
	175	181	180	176	185	187	184	(194)				
P. variotii	0.881	1.158	1.188	1.158	0.841	0.833						
	180	184	180	179	187	187	(194)					
T. harzianum	0.871	1.111	1.023	1.108	0.691							
	182	184	179	173	190	(190)						
T. reesei II	0.895	1.295	1.227	1.306								
	183	181	179	180	(190)							
D. thermophilum	1.126	0.913	1.015									
	180	198	197	(199)								
B. subtilis B230	0.854	0.077										
	179	204	(205)									
B. agaradhaerens	1.051											
	181	(207)										
Streptomyces sp. S38	(185)											

Table 6
Salt bridges (5 Å cutoff) in thermostable xylanases.

Salt bridges found in all thermophilic xylanase structures are highlighted in bold. Where the positive and negative residues of the salt bridge are reversed, they are indicated in parentheses. Where an equivalent pair exists in a thermostable xylanase but the minimum interatomic distance of the charged atoms is greater than 5 Å, they are indicated with a > sign.

B. subtilis B230	*D. thermophilum*	*P. variotii*	*T. lanuginosus*
Asp13>His12	—	Asp11–His10	Asp11–His10
Glu18–Arg49	Glu16–Arg47	—	—
—	—	—	Glu31–H10
—	—	Glu36–Arg187	Glu36–Arg187
—	—	—	Asp63–His61
—	—	Arg81–Asp129	Arg81–Asp129
Glu94–Arg128	Glu90–Arg125	Glu86>Arg122	Glu86>Arg122
Asp99–Arg151	**Glu95–Arg148**	**Glu91–Arg145**	**Glu91–Arg145**
Asp99–Arg147	**Glu95–Arg144**	**Glu91–Arg141**	**Glu91–Arg141**
Asp122–Arg105	Asp119–Arg101	—	(Arg116–Asp97)
Glu125–Lys111	—	—	—
Glu125–Lys141	(Arg122–Asp138)	—	(Lys119–Asp135)
Asp117–Arg151	**Asp114–Arg148**	**Asp111–Arg145**	**Asp111–Arg145**
Asp117–His161	**Asp114–His158**	**Asp111–His155**	**Asp111–His155**
—	Asp173–Lys51	—	—
Glu177–Arg49	—	—	—
Glu177–Lys53	—	—	—
—	—	Asp192–Arg58	Asp192–Arg58

low estimate for β-strand content can be explained by the fact that xylanase contains β-sheets that are highly curved; since these β-sheets show large variations in their φ, ψ angles, this may have resulted in outlying behaviour in the CD spectra.

4. Discussion

4.1. The *B. subtilis* B230 xylanase structure

Xylanase from *B. subtilis* B230 adopts the canonical family 11 xylanase fold, with two closely packed β-sheets in a jelly-roll configuration. The smaller of the two, sheet *B*, contains all the active-site residues (Fig. 2*a*). The strands that make up the sheets are numbered sequentially by position in sheet (*A*1–6 and *B*1–9; Fig. 3). A single α-helix packs against sheet *B*. Sheet *B* forms an elongated open cleft that allows a xylan polymer to bind with extensions on either side of the scissile xylopyranoside residues (Sabini *et al.*, 1999).

By homology with other family 11 xylanases, the key catalytic residues of Xyn11X are Glu94 and Glu183 (Fig. 2). Miao *et al.* (1994) showed that Glu78 in *B. subtilis* family 11 xylanase (equivalent to Glu94 in Xyn11X) acts as the nucleophile in the enzymatic reaction. Site-directed mutagenesis of Glu78 and Glu172 in *B. circulans* (equivalent to Glu94 and Glu183 in Xyn11X) has confirmed that these residues are essential for catalysis (Wakarchuk, Campbell *et al.*, 1994). Tyr80 and Tyr69 in *B. circulans* xylanase (equivalent to Tyr96 and Tyr85 in Xyn11X) have been shown to be crucial for catalytic activity in that the Y80F mutant shows little activity and the Y69F mutant is inactive (Wakarchuk, Campbell *et al.*, 1994; Sidhu *et al.*, 1999). These and other residues (including Gln142) are thought to lower the pK_a of Glu94, ensuring that it remains charged and hence a nucleophile. The measured pK_a of GluE78 in *B. circulans* xylanase is 4.6 (McIntosh *et al.*, 1996). The C^{β} and C^{γ} atoms of Glu94 are part of a contiguous cluster of hydrophobic residues in the active site of Xyn11X (Fig. 2*b*). Tyr96 and Tyr85 form part of this cluster, as does Trp87, which is absolutely conserved in all solved xylanase structures. Thus, reducing the flexibility of Glu94 seems to be important for xylanase activity. Any effort to increase the thermostability of Xyn11X should not involve altering these residues.

The high degree of structural similarity of the xylanase structures from divergent organisms (Fig. 3, Table 5) may indicate that they are optimal for enzymatic function and have thus been conserved throughout evolution. When the regions of structural similarity are mapped onto the Xyn11X structure (data not shown), it is apparent that the shape of the active-site cleft is highly conserved. Thus, in an effort to improve structural stability, these regions should not be altered. The regions that do not conform structurally are mostly in sheet *A*, which contributes no residues to the active site, and in the region of strands *B*1 and *B*2, which are not close to the catalytic residues. The β-strands and α-helices in Xyn11X are found in all other xylanases, with the exception of strands *A*1 and *B*1, which are not present in four of the aligned xylanases

(*T. reesei* isozyme I, *A. kawachii*, *A. niger* and *B. circulans*; Fig. 3) owing to a shorter N-terminal region. These strands appear to bury hydrophobic residues in the core of the protein and thus may contribute to thermostability in the xylanases that possess them.

4.2. Phylogenetic analysis

Examination of the phylogenetic tree constructed for known family 11 xylanase sequences (Fig. 4) shows that the enzymes for which there is direct evidence for thermostability fall into three distinct clades. The thermostable isozymes from *B. subtilis* B230, *D. thermophilum* and a number of *Clostridium* species occur in clade 1. Xylanases from *P. variotii* and *T. lanuginosus* are closely related and occur in clade 2. Clade 3 contains xylanases from the moderately thermotolerant species *Streptomyces thermoviolaceus*, *S. thermocyaneoviola* and *B. circulans* and the xylanases from the thermophiles *B. stearothermophilus* and *Thermomonospora fusca*. The phylogenetic analysis presented here supports an earlier suggestion that family 11 xylanase fold may be prone to giving rise to thermostable proteins (Fontes *et al.*, 1995). The finding of closely related bacterial and fungal enzymes within clades supports previous suggestions (Gilbert *et al.*, 1992) of a prokaryotic origin (possibly from the rumen bacterium *Ruminococcus flavefaciens*) of the fungal xylanases of the rumen fungus *Neocallimastix patriciarum*. The analysis presented here suggests that the rumen fungus *Piromyces communis* has also obtained the xylanase gene *via* horizontal gene transfer from rumen prokaryotes.

Given the presence of thermostable xylanases in multiple distinct clades, it is expected that different adaptations are found in distinct subfamilies of this enzyme. It is unlikely that all adaptations contributing to thermostability are present in a single isoenzyme. With this in mind, adaptations of the xylanases contributing to thermostability are discussed.

Table 7
Structural data derived from solved xylanases.

PDB files used for the calculations are given in Fig. 3.

	Hydrogen bonds†	%Crg‡	%Pol§	pI¶
T. reesei I	157 (22)	8.3	75.6	5.00
A. kawachii	177 (38)	17.3	82.5	3.88
A. niger	170 (38)	17.6	82.5	3.99
B. circulans	180 (36)	18.3	78.4	9.05
T. lanuginosus	199 (39)	26.8	71.8	4.69
P. variotii	196 (38)	18.6	73.7	4.78
T. harzianum	188 (32)	11.9	73.5	8.71
T. reesei II	183 (30)	12.1	72.9	8.73
D. thermophilum	205 (42)	17.8	82.0	8.76
B. subtilis B230	210 (42)	22.8	75.8	8.83
B. agaradhaerens	202 (36)	21.1	76.2	8.58
Streptomyces sp. S38	176 (25)	16.0	62.9	8.58

† The number of side chain to side chain hydrogen bonds is given in parentheses. ‡ The percentage of accessible surface area composed of charged residues. § The percentage of accessible surface area composed of polar residues. ¶ The computed pI.

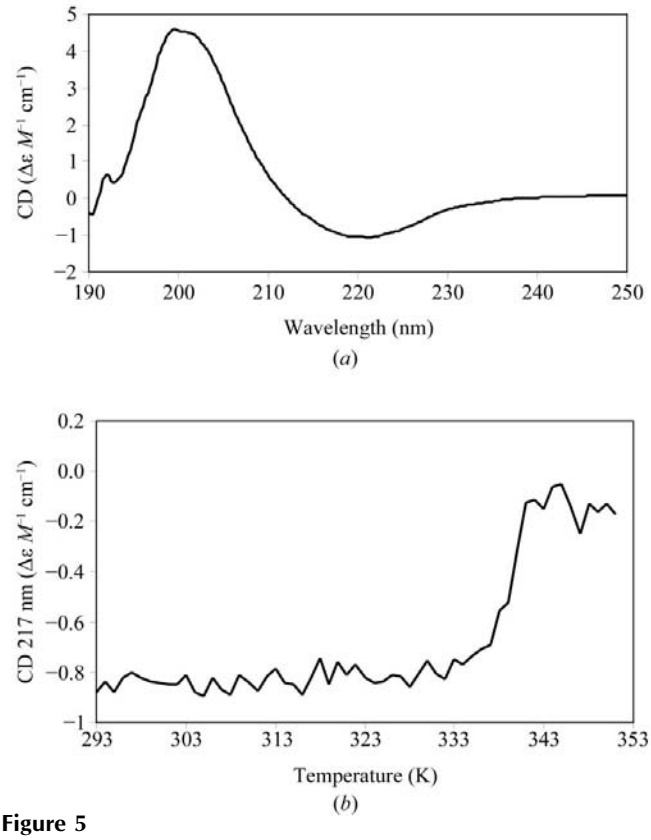

Figure 5
(*a*) CD spectra of Xyn11X. (*b*) CD at 217 nm measured at 1 K intervals.

4.3. Thermostability

Since structural and whole-genome data for thermophiles and corresponding proteins from mesophiles became available, several investigators have analysed the data in the hope of discovering the determinants of thermal stability in proteins. Based on whole-genome analyses, Das & Gerstein (2000) concluded that thermostable proteins have a higher proportion of charged amino acids compared with mesophiles. They also concluded that the charged residues in thermophiles were arranged such that the dipole of intrahelical charged pairs tended to align with the helix dipole. Based on structural comparisons, Kumar *et al.* (2000) found that the number of salt bridges and side chain to side chain hydrogen bonds increase in thermostable proteins. On examination of the salt bridges of the thermostable xylanases, it becomes apparent that the enzymes have evolved different patterns, although there are some salt bridges that are preserved across all structures in the family (Table 6). Several pairs are found in the more closely related Xyn11X and *D. thermophilum* xylanases but not in *P. variotii* and *T. lanuginosus* xylanases and *vice versa*. In a number of cases, the positive- and negative-charge contributors to the bridge are swapped with regard to their positions in the structures, suggesting that the salt bridge is important for thermostability. One such pair links the *B*7 and *B*8 strands (Glu125–Lys141 in Xyn11X and Lys119–Asp135 in *T. lanuginosus* xylanase). These two strands form a lid over the active site and are linked by the 'thumb' motif (Fig. 2). Another such bridge links strand *B*8 with the 'cord' between strands *B*6 and *B*9. In Xyn11X the pair is Asp122–Arg105 and in *T. lanugi-*

nosus it is Arg116–Asp97. Again, the elements of secondary structure linked are close to the active site.

Analysis of side chain to side chain hydrogen bonds (Table 7) shows that the thermostable xylanases from *D. thermophilum* and *B. subtilis* B230 have the most hydrogen bonds of this kind, in accordance with the observations of Kumar *et al.* (2000). The xylanases with the next largest number of side chain to side chain hydrogen bonds are the *B. agaradhaerens* and the thermostable *T. lanuginosus* and *P. variotii* xylanases. When the proportions of charged and polar surface residues are examined, the data are not so clear. The xylanase with the highest proportion of surface area made up by charged amino acids is the *T. lanuginosus* isozyme; however, other thermostable xylanases have proportions similar to the ostensibly mesophilic xylanases. With regard to the percentage of surface residues that are polar, the mesophilic xylanases *A. kawachii* and *A. niger* have the highest ratios.

D. thermophilum xylanase has a short 3_{10}-helix between strands *B*3 and *A*5 that appears to prevent the exposure to solvent of hydrophobic clusters sandwiched between the two sheets (McCarthy *et al.*, 2000). A 3_{10}-helix is also present in *B. agaradhaerens* xylanase, which may be thermostable although it has not been tested as such. This helix is absent in Xyn11X, although a loop in a similar conformation is present. The loop contributes hydrophobic residues to the cluster and possesses polar residues exposed to the medium. A similar 3_{10}-helix could feasibly be introduced into Xyn11X by site-directed mutagenesis.

The aforementioned loop and helix are not present in the thermostable xylanases from *P. variotii* (Kumar *et al.*, 2000) and *T. lanuginosus* (Gruber *et al.*, 1998). However, these enzymes possess stabilizing disulfide bridges connecting strand *B*9 and the α-helix that are not found in any other xylanase crystal structures.

An aromatic ring-stacking interaction (identified as Trp8 and Phe14 in *T. lanuginosus* xylanase) has been shown to be important for thermostability in those xylanases which possess aromatic residues in equivalent positions (Georis *et al.*, 2000). Trp8 and Phe14 are located on strands *B*1 and *B*2, respectively. Xyn11X and some other thermostable xylanases lack aromatic residues at these locations.

A feature that appears to be critical for the stabilization of the xylanase is a cluster of buried charged and polar residues surrounding a water molecule. This water interacts with Asp117 at the end of strand *B*9, Tyr121 in strand *B*8 and His161 in the α-helix. This water has been found in all xylanases so far studied and these residues are conserved across the xylanases (Fig. 3). Together with this cluster are salt bridges between Asp117 and Arg151 and between Asp99, Arg147 and Arg151. Asp99 is replaced by glutamate in some xylanases, but the essential interactions are maintained in all structures.

Efforts to improve the thermostability of family 11 xylanases have been partially successful. Wakarchuk, Sung *et al.* (1994) engineered disulfide bridges into the *B. circulans* xylanase between the α-helix and strand *B*9 *via* site-directed mutagenesis. While all of the disulfide bonds tested increased the thermostability of the *B. circulans* xylanase, not all enhanced the activity of the enzyme at elevated temperatures. Georis *et al.* (2000) were able to improve the thermostability and thermophilicity of the mesophilic family 11 xylanase from *Streptomyces* sp. S38 by site-directed mutagenesis. They created an aromatic interaction involving residues on strands *B*1 and *B*2 (Tyr11–Tyr16), increasing the stability of the N-terminal part of the structure.

4.4. Conclusion

Not all features identified as contributing to the stability of family 11 xylanases are present in Xyn11X. Thus, to improve the thermostability of Xyn11X, additional salt bridges as found in other thermostable xylanases could be engineered into its sequence. Furthermore, the site-directed mutant T11Y.D16Y would create an aromatic interaction cluster similar to that observed to improve thermostability of *Streptomyces* sp. S38 xylanase (Georis *et al.*, 2000). Finally, disulfide bonds similar to those introduced into the *B. circulans* xylanase sequence (Wakarchuk, Sung *et al.*, 1994) could be engineered. Further changes to Xyn11X might involve mutations that lead to the formation of additional hydrophobic interactions on the surface of the enzyme.

The authors would like to thank Robert Dunlop for helpful comments in the preparation of this manuscript. We would also like to thank Roger Latham for assistance with protein purification.

References

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

Campbell, R. L., Roes, D. R., Wakarchuk, R. J., To, R. J., Sung, W. & Yaguchi, M. (1993). *Proceedings of the Second TRICEL Symposium on Trichoderma reesei Cellulases and other Hydrolases*, edited by P. Suominen & T. Reinikainen, pp. 63–72. Helsinki: Foundation for Biotechnical and Industrial Fermentation Research.

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

Crennell, S., Garman, E., Laver, G., Vimr, E. & Taylor, G. (1994). *Structure*, **2**, 535–544.

Das, R. & Gerstein, M. (2000). *Funct. Integr. Genomics*, **1**, 76–88.

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

Delbaere, L. T., Vandonselaar, M., Prasad, L., Quail, J. W., Wilson, K. S. & Dauter, Z. (1993). *J. Mol. Biol.* **230**, 950–965.

Dunlop, R. W., Wang, B., Ball, D., Ruollo, A. B. & Falk, C. J. (1996). Australian Patent 695 276.

Fontes, C. M. G. A., Hazlewood, G. P., Morag, E., Hall, J., Hirst, B. H. & Gilbert, H. J. (1995). *Biochem. J.* **307**, 151–158.

Fukusaki, E., Panbangred, W., Shinmyo, A. & Okada, H. (1984). *FEBS Lett.* **171**, 197–201.

Fushinobu, S., Ito, K., Konno, M., Wakagi, T. & Matsuzawa, H. (1998). *Protein Eng.* **11**, 1121–1128.

Gebler, J., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Beguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Warren, R. A. J. & Withers, S. G. (1992). *J. Biol. Chem.* **267**, 12559–12561.

Georis, J., de Lemos Esteves, F., Lamotte-Brasseur, J., Bougnet, V., Devreese, B., Giannotta, F., Granier, B. & Frere, J. M. (2000). *Protein Sci.* **9**, 466–475.
Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W. & Kratky, C. (1998). *Biochemistry*, **37**, 13475–13485.
Henrissat, B., Teeri, T. T. & Warren, A. (1998). *FEBS Lett.* **425**, 352–354.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Joshi, M. D., Sidhu, G., Pot, I., Brayer, G. D., Withers, S. G., McIntosh, L. P. (2000). *J. Mol. Biol.* **299**, 255–279.
Kabsch, W. & Sander, C. (1983) *Biopolymers*, **22**, 2577–2637.
Kleywegt, G. J. (1996) *Acta Cryst.* D**52**, 842–857.
Krengel, U. & Dijkstra, B. W. (1996) *J. Mol. Biol.* **263**, 70–78.
Kulkarni, N., Shendye, A. & Rao, M. (1999). *FEMS Microbiol. Rev.* **23**, 411–456.
Kumar, P. R., Eswaramoorthy, S., Vithayathil, P. J. & Viswamitra, M. A. (2000). *J. Mol. Biol.* **295**, 581–593.
Kumar, S., Tsai, C.-J. & Nussinov, R. (2000). *Protein Eng.* **13**, 179–191.
McCarter, J. D. & Withers, S. G. (1994). *Curr. Opin. Struct. Biol.* **4**, 885–892.
McCarthy, A. A., Morris, D. D., Bergquist, P. L. & Baker, E. N. (2000). *Acta Cryst.* D**56**, 1367–1375.
McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W. & Withers, S. G. (1996). *Biochemistry*, **35**, 9958–9966.
Miao, S., Ziser, L., Aebersold, R. & Withers, S. G. (1994). *Biochemistry*, **33**, 7027–7032.
Neilson, A. H., Allard, A. S., Hynning, P. A. & Remberger, M. (1991). *Toxicol. Environ. Chem.* **30**, 3–41.
Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Russell, R. B. & Barton, G. S. (1992). *Proteins Struct. Funct. Genet.* **14**, 309–323.
Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jorgensen, P. L., Schulein, M., Dupont, C., Davies, G. J. & Wilson, K. S. (1999). *Chem. Biol.* **6**, 483–492.
Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L. & Brayer, G. D. (1999). *Biochemistry*, **38**, 5346–5354.
Sreerama, N. & Woody, R. W. (2000). *Anal. Biochem.* **287**, 252–260.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Törrönen, A., Harkki, A. & Rouvinen, J. (1994). *EMBO J.* **13**, 2493–2501.
Törrönen, A. & Rouvinen, J. (1995). *Biochemistry*, **34**, 847–856.
Törrönen, A. & Rouvinen, J. (1997). *J. Biotechnol.* **57**, 137–149.
Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.
Viikari, L., Kantelinen, A., Sundquist, J. & Linko, M. (1994). *FEMS Microbiol. Rev.* **13**, 335–350.
Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J. & Yaguchi, M. (1994). *Protein Sci.* **3**, 467–475.
Wakarchuk, W. W., Sung, W. L., Campbell, R. L., Cunningham, A., Watson, D. C. & Yaguchi, M. (1994). *Protein Eng.* **7**, 1379–1386.
Wouters, J., Georis, J., Engher, D., Vandenhaute, J., Dusart, J., Frere, J. M., Depiereux, E., Charlier, P. (2001). *Acta Cryst.* D**57**, 1813–1819.