Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of XynB, a highly thermostable β-1,4-xylanase from *Dictyoglomus thermophilum* Rt46B.1, at 1.8 Å resolution

Andrew A. McCarthy,[a] Daniel D. Morris,[b] Peter L. Bergquist[b,c] and Edward N. Baker[a]*

[a]School of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland, New Zealand, [b]Department of Biological Sciences, Macquarie University, NSW 2109, Australia, and [c]Department of Molecular Medicine, University of Auckland Medical School, Private Bag 92-019, Auckland, New Zealand

Correspondence e-mail: ted.baker@auckland.ac.nz

Received 8 May 2000
Accepted 14 July 2000

PDB Reference: XynB, 1f5j.

Microorganisms employ a large array of enzymes to break down the cellulose and hemicelluloses of plant biomass. These enzymes, especially those with high thermal stability, have many uses in biotechnology. We have solved the crystal structure of a β-1,4-xylanase, XynB, from the extremely thermophilic bacterium *Dictyoglomus thermophilum*, isolate Rt46B.1. The protein crystallized from 1.6 *M* ammonium sulfate, 0.2 *M* HEPES pH 7.2 and 10% glycerol, with unit-cell parameters $a = b = 91.3$, $c = 44.9$ Å and space group $P4_3$. The structure was solved at high resolution (1.8 Å) by X-ray crystallography, using the method of isomorphous replacement with a single mercury derivative, and refined to a final R factor of 18.3% ($R_{\rm free}$ = 22.1%). XynB has the single-domain fold typical of family 11 xylanases, comprising a jelly roll of two highly twisted β-sheets that create a deep substrate-binding cleft. The two catalytic residues, Glu90 and Glu180, occupy this cleft. Compared with other family 11 xylanases, XynB has a greater proportion of polar surface and has a slightly extended C-terminus that, combined with the extension of β-strand *A*5, gives additional hydrogen bonding and hydrophobic packing. These factors may account for the enhanced thermal stability of the enzyme.

1. Introduction

Cellulose and hemicelluloses are complex polysaccharides, the main components of plant biomass and the largest source of carbon in Nature. Microbes have evolved to utilize this carbon source and secrete many different specific enzymes which function synergistically in the breakdown of cellulose and hemicelluloses. These glycosyl hydrolases can be classified into a total of 45 families, including 11 1,4-glycosidase families, on the basis of sequence similarity (Henrissat, 1991; Henrissat & Bairoch, 1993). Many three-dimensional structures of glycosyl hydrolases have been determined, revealing different folds and significant variations in the active-site construction among enzymes from different families (Davies & Henrissat, 1995).

Xylan is a hemicellulose composed of a 1,4-β-linked xylopyranoside backbone with varied branched heteropolysaccharides depending on its botanical origin. The enzymes acting on the backbone of xylans are endo-1,4-β-xylanases (E.C. 3.2.1.8), capable of hydrolysing the internal 1,4-β-bonds of xylan, and exo-1,4-β-xylosidases (E.C. 3.2.1.37), which prefer to hydrolyse the 1,4-glycosidic linkages from the ends of the polymeric chain. The enzymatic hydrolysis of xylan has drawn commercial interest because of potential industrial applications, for example in animal feeds and in the pulp-bleaching process, where xylanases hold the

© 2000 International Union of Crystallography
Printed in Denmark – all rights reserved

promise of reducing the amount of chlorine toxins released (Viikari *et al.*, 1994). Thus, research in this area has focused both on identifying new xylanases and improving the thermostability of others to withstand the harsh environment involved in pulp bio-bleaching or in the pelleting process for animal feeds.

The xylanases that have been characterized to date fall into two structural families, F and G, corresponding to families 10 and 11 in the numerical classification of glycosyl hydrolases. Family 10 xylanases have a molecular mass of about 35 kDa and have been shown to have the $(\alpha/\beta)_8$ or TIM-barrel fold (Davies & Henrissat, 1995; Harris *et al.*, 1994). Family 11 xylanases are smaller, about 20 kDa, with a single-domain structure based on two highly twisted β-sheets that pack against each other and curve around to create a large substrate-binding cleft. The overall shape has been compared to the shape of a right hand (Törrönen *et al.*, 1994) and has the same 'jelly-roll' β-barrel topology as the plant lectins (Keitel *et al.*, 1993).

Glycosidases may have two alternative enzymatic mechanisms of glycosidic bond cleavage, depending on whether there is a net retention or inversion of the anomeric configuration (McCarter & Withers, 1994). The family 11 xylanases have a retaining mechanism (Gebler *et al.*, 1992) and contain a pair of carboxylic acids at the active site. One, with the lower pK_a, functions as a nucleophile, whereas the other acts as a general acid/base in a double-displacement mechanism (McCarter & Withers, 1994) and has been shown to cycle over a pK_a range of 2.5 during catalysis (McIntosh *et al.*, 1996). A sequence-conserved arginine in the vicinity of the active site has been proposed to lower the pK_a of the nucleophile (Wakarchuk, Campbell *et al.*, 1994). The family 11 xylanases can be further classified in terms of their pH optimum. A mutation of an asparagine to an aspartate at the active site is believed to account for the shift to a lower pH optimum (Krengel & Dijkstra, 1996).

Three-dimensional structures have been determined for the catalytic domains of a number of family 11 xylanases from both fungi and bacteria. These include enzymes from *Bacillus circulans* (Campbell *et al.*, 1993; Wakarchuk, Campbell *et al.*, 1994; Sidhu *et al.*, 1999), *Trichoderma harzianum* (Campbell *et al.*, 1993), *Trichoderma reesei* (Törrönen *et al.*, 1994; Törrönen & Rouvinen, 1995), *Aspergillus niger* (Krengel & Dijkstra, 1996), *Bacillus agaradhaerens* (Sabini *et al.*, 1999), *Paecilomyces varioti* Bainier (Kumar *et al.*, 2000) and *Thermomyces lanuginosus* (Gruber *et al.*, 1998). These enzymes differ in detail in their enzymatic properties and in their thermal stability. Notably, two of them, the enzymes from *P. varioti* and *T. lanuginosus*, are thermophilic, the latter having an optimum temperature for activity of 343 K.

Here, we present the three-dimensional structure of a family 11 xylanase from *Dictyoglomus thermophilum* isolate Rt46B.1 determined by X-ray crystallography at 1.8 Å resolution. *D. thermophilum* Rt46B.1 is an extremely thermophilic strictly anaerobic bacterium that produces both family 10 and family 11 xylanases that act over a broad pH range (pH 5.5–9.0) at 353 K. Several variants of the family 11 xylanase from this organism have been cloned and expressed in *Escherichia coli* (Morris *et al.*, 1998). The variant we have analysed is XynB3; this variant has 199 amino-acid residues and the N-terminal sequence ALTSNA. It lacks the normal leader sequence, whose removal gave higher expression levels in *E. coli* and it also lacks the C-terminal non-catalytic domain which reduces thermostability (Morris *et al.*, 1998). XynB3 has a pH optimum of 6.5 and a temperature optimum of 348 K and retains 90% of its activity after heating at 353 K for 8 h. This is the most thermostable xylanase yet to be characterized structurally and allows us to draw comparisons between mesophilic and thermophilic xylanases of this family.

2. Materials and methods

2.1. Crystallization and data collection

Expression and purification of the β-1,4-xylanase XynB from *D. thermophilum* was carried out as described previously (Morris *et al.*, 1998). Good-quality crystals were obtained at 291 K by vapour diffusion in hanging drops from a solution containing 10 mg ml^{-1} protein, 1.5–1.6 *M* ammonium sulfate, 0.2 *M* HEPES buffer at pH 7.2 and 10% glycerol. The crystals, which required approximately two weeks to grow to maximal size, were tetragonal, space group $P4_3$, with two molecules in the asymmetric unit and a solvent content of approximately 45%. Room-temperature data from native and derivative crystals were collected with a MAR Research MAR345 imaging-plate system mounted on a Rigaku RU-300 X-ray generator operating at 50 kV and 100 mA equipped with a copper anode (λ = 1.54 Å) and Franck's double focusing mirrors. A second native data set to 1.8 Å was also collected at 110 K, using synchrotron radiation (wavelength 1.08 Å) on beamline 7.1 at the Stanford Synchrotron Radiation Laboratory; the crystal was flash-frozen after being soaked for 30 s in a cryo-protectant consisting of 1.7 *M* ammonium sulfate, 0.2 *M* HEPES buffer pH 7.2 and 20% glycerol. All data were processed and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Table 1 contains a summary of the data collection and processing.

2.2. Phasing

The structure was solved using the SIRAS method. A single mercury derivative was prepared by soaking crystals in standard mother liquor (1.7 *M* ammonium sulfate, 0.2 *M* HEPES buffer pH 7.2 and 10% glycerol) containing 50 m*M* $Hg(CH_3COO)_2$. The isomorphous difference Patterson map, calculated from native and derivative room-temperature data, was solved both manually and using the Patterson heavy-atom search method implemented in *CNS* (Brunger *et al.*, 1998). Only two sites could be located and no further sites were found in difference Fourier maps. These sites and their enantiomorph were refined and phases were calculated using *MLPHARE* (Otwinowski, 1991) from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). Choice of the correct enantiomorph was made by inspecting the $R_{\rm free}$ output on solvent flattening with *DM* (Cowtan &

Table 1
Data-collection, data-processing and phasing statistics.

Data set	Native (110 K)	Native (293 K)	$Hg(CH_3COO)_2$
Source	Synchrotron	Rotating anode	Rotating anode
Wavelength (Å)	1.08	Cu $K\alpha$ 1.54	Cu $K\alpha$ 1.54
Unit-cell parameters			
$a = b$ (Å)	91.29	91.98	92.0
c (Å)	44.93	45.83	45.79
$\alpha = \beta = \gamma$ (°)	90	90	90
Resolution (Å)	1.8	2.0	2.3
No of reflections measured	207114	80713	101007
No. of unique reflections	34176	25173	17034
Data completeness (%)	99 (97)	97 (97)	99 (98)
R_{merge} (%)	8.9 (30.7)	5.6 (13.1)	8.5 (19.5)
R_{iso} (%)			22.4
No. of heavy-atom sites			2
R_{cullis} centric/acentric (ano)			0.46/0.32 (0.86)
Phasing power centric/acentric (ano)			2.3/5.4 (1.54)
Mean FOM after SIRAS centric/acentric			0.40/0.57

Main, 1998). Phases were further improved by *SHARP* (de La Fortelle & Bricogne, 1997) and solvent flattened using *DM* (Cowtan & Main, 1998) with 35% solvent content to give an overall figure of merit of 0.49 to 2.3 Å resolution.

2.3. Phase extension and improvement using *wARP*

The program *wARP* (Perrakis *et al.*, 1997; van Asselt *et al.*, 1998) was used for accelerated structure elucidation. The solvent-flattened phases to 2.3 Å were improved and extended to 2.0 Å using the procedures in *wARP*; automatic tracing and model building into the electron-density map located 260 residues. This initial model allowed the identification of the NCS operator. The initial phases from *SHARP* (de La Fortelle & Bricogne, 1997) were then solvent flattened and averaged using *DM* (Cowtan & Main, 1998). The new phases were again input into *wARP* (Perrakis *et al.*, 1997) as described above and resulted in an initial model of 170 residues for one of the molecules in the asymmetric unit.

2.4. Model building and refinement

Further model building was performed with the interactive graphics program *O* (Jones *et al.*, 1991). The quality of the *wARP* (Perrakis *et al.*, 1997) map and the initial model allowed the remainder of the structure to be built without difficulty. This initial room-temperature model was then positioned in the 110 K cell with *AMoRe* (Navaza, 1994) and the 1.8 Å synchrotron data set was used for all subsequent refinement. Refinement was carried out using the maximum-likelihood target with bulk-solvent correction in *REFMAC* (Murshudov *et al.*, 1997) and medium main-chain/side-chain NCS restraints. The refinement was guided by monitoring the free R factor calculated with 10% of the data excluded from the refinement (Brünger, 1992). Potential solvent molecules were located with the building of the solvent structure method in *wARP* (Perrakis *et al.*, 1997). All solvent molecules were refined with unit occupancy. In addition, the solvent molecules were visually inspected and were deleted from the model if they were not within hydrogen-bonding distance of one or more potential hydrogen-bond partners or were close to regions in which the protein density was ambiguous. The final refinement was performed with *CNS* (Brunger *et al.*, 1998) with a bulk-solvent correction. Refinement details are summarized in Table 2.

Table 2
Refinement details.

Resolution range (Å)	500–1.8
R_{cryst} (%) (No. of reflections)	18.3 (30705)
R_{free} (%) (No. of reflections)	22.1 (3404)
No. of atoms (non-H)	
Total	3459
Protein	3154
Water (ions)	295 (10)
Deviations from ideal geometry	
Bond lengths (Å)	0.006
Angles (°)	1.5
Average B factors (Å^2)	
All atoms	12.9
Protein atoms	11.9
Solvent atoms (ions)	22.5

2.5. Quality of the final model

The final model comprises all 199 residues for both molecules (labelled *A* and *B*) in the asymmetric unit, together with 295 water molecules and two sulfate ions. This model gives an R factor of 18.3% for all data to 1.8 Å resolution (no σ-cutoff) and a free R (for 10% of data) of 22.1%. The protein molecules conform closely to standard bond lengths and angles as defined by Engh & Huber (1991), with root-mean-square (r.m.s.) deviations of 0.006 Å and 1.5°, respectively. A Ramachandran plot for the main-chain dihedral angles showed that 87.9% of all the residues (excluding glycine) have dihedral angles that fall in the most favoured regions as defined by *PROCHECK* (Laskowski *et al.*, 1993). There are no outliers. Two *cis*-proline residues are found in the structure, Pro87 and Pro103. Residues Thr21*A*, Ser35*A*, Ser67*A*, Ser150*A*, Thr21*B*, Thr69*B*, Ser146*B* and Ser150*B* were modelled in two alternate side-chain conformations. The Tyr52–Asn53 peptide bond was also modelled in two alternate conformations for molecule *B*, though not for molecule *A*, which is stabilized by crystal packing contacts in this region.

2.6. Analysis of the model

Hydrogen bonds in the structures of XynB and other family 11 xylanases were identified using the criteria of Baker & Hubbard (1984), after first checking Asn, Gln and His residues for the correct side-chain orientation (McDonald & Thornton, 1994). Ion pairs were inferred where the distance between potentially oppositely charged atoms was less than 4.0 Å (Barlow & Thornton, 1983), but with the following proviso. Where two hydrogen bonds were made between partners in an ionic interaction (*e.g.* between N^{ε} and $N^{\eta 2}$ atoms of arginine and the $O^{\varepsilon 1}$ and $O^{\varepsilon 2}$ atoms of glutamate) this was treated as

two ion pairs, whereas when only a single hydrogen bond in an ion-pair interaction was possible it was treated as one. Calculations of solvent-accessible surface were made with *AREAIMOL* (Collaborative Computational Project, Number 4, 1994) using a 1.4 Å radius probe. Multiple structural superpositions were performed with the program *MAPS* (Lu, 1998). This method uses the algorithm *TOP* (Lu, 1996) to perform a structural alignment for each pair of proteins and calculates pairwise structural diversity scores which are root-mean-square differences in C^{α} positions normalized by the number of atom equivalences relative to the average number in the family.

3. Results and discussion

3.1. Overall structure

XynB from *D. thermophilum* Rt46B.1 is folded into a single domain based on two β-sheets that pack against one another. The smaller sheet *A* comprises five antiparallel strands, *A*2–*A*6, and the larger sheet *B* comprises nine mostly antiparallel strands, *B*1–*B*9. Sheet *B* curves around to create a large cleft where the xylan substrate binds and the catalytic pair Glu90 and Glu180 are located. A single α-helix, residues 155–166 in XynB, packs against sheet *B* at the back of the molecule. This fold, shown in Fig. 1, is characteristic of the family 11 xylanases and has been likened by Törrönen *et al.* (1994) to a hand in which the two β-sheets and the α-helix form the 'fingers' and 'palm' and a loop at the end of the extended strands B7 and B8 forms a 'thumb' that helps to enclose the active-site cleft. The main structural deviation seen in this family is that some members, the xylanase from *B. circulans* (Campbell *et al.*, 1993), xylanase I from *Tr. reesei* (Törrönen & Rouvinen, 1995) and xylanase I from *A. niger* (Krengel & Dijkstra, 1996), lack the first strand *B*1 of sheet *B* and are thus slightly smaller in size.

A multiple sequence alignment of XynB with six other family 11 xylanases, including the thermophilic xylanases from *P. varioti* and *T. lanuginosus*, is given in Fig. 2. All of these enzymes share substantial sequence identity that ranges from 46 to 62% pairwise identity, except for the *P. varioti*/*T. lanuginosus* and *Tr. harzianum*/*Tr. reesei* II pairs, which have around 90% identity (Table 3). We have further used the program *MAPS* (see §2.6) to carry out a multiple structure alignment of XynB with these other family 11 xylanases. All of them can be aligned structurally over nearly 90% of C^{α} atoms with a 3.0 Å cutoff. Root-mean-square differences in C^{α} positions between XynB and the other enzymes are all in the range 1.0–1.2 Å. When the structure diversity scores calculated by *MAPS* are compared, however, it is clear that XynB groups most closely with the xylanases from *P. varioti, T. lanuginosus* (both thermostable) and *Tr. harzianum*.

The largest difference between XynB and the other family 11 xylanase structures occurs in the loop connecting strand *B*3 to strand *A*5. An insertion of five extra residues in this loop in XynB causes residues 52–61 to take on a different conformation; residues 55–59 form a 3_{10}-helix and the loop as a whole folds back against the body of the molecule, burying several hydrophobic residues including Tyr52, Trp56, Leu59 and Ile62. The β-sheet hydrogen bonding between strands *A*5 and *A*4 (the C-terminal strand) is also extended, as the *B*3–*A*5

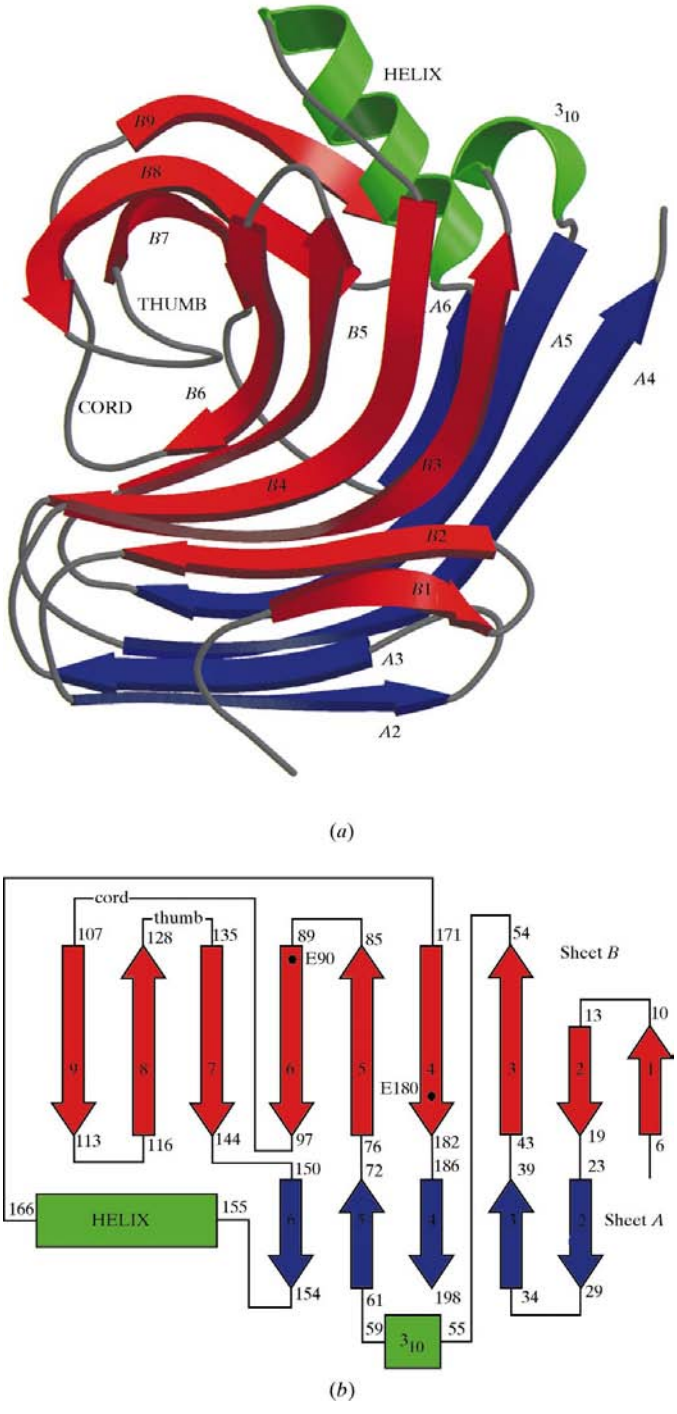

Figure 1
(*a*) Ribbon diagram for XynB, the xylanase from *D. thermophilum* Rt46B.1, showing the overall fold. The two β-sheets are shown in blue (sheet *A*) and red (sheet *B*), with the β-strands and other structural features labelled as in Törrönen *et al.* (1994). Figure drawn with *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Murphy, 1994). (*b*) Topology diagram for XynB, showing the residue numbers for the secondary-structural elements and the location of the catalytically essential residues Glu90 and Glu180.

insertion is accompanied by an extension of the C-terminus; the latter is six residues longer in XynB than in the *B. circulans* and *Tr. harzianum* xylanases and three residues longer than in the *P. varioti* and *T. lanuginosus* enzymes. A further difference is in the extended nine-residue loop, 98–106, that closes the cleft on one side, described as a 'cord' by Törrönen *et al.* (1994). This loop is one residue shorter in XynB and also contains a *cis*-proline at Pro103.

3.2. Active site

By analogy with the other family 11 xylanases, the two catalytic acid groups in XynB are provided by Glu90 and Glu180. For glycosidases with a retaining mechanism, the distance between the carboxyl groups (defined as the mean separation of the carboxylic atoms) is usually around 5.5 Å. The distance between Glu90 and Glu180 in XynB is 6.5 Å, similar to the potential 'active conformation' observed in the xylanases from *B. circulans* (5.9 Å), *Tr. reesei* I (6.7 Å) and *Tr. harzianum* (6.4 Å) and shorter than for the *T. lanuginosus* (10.7 Å) and *Tr. reesei* II (10.9 Å) enzymes; in the latter, a conformational change is expected to generate an active species. We conclude, therefore, that Glu90 and Glu180 in XynB are in their active conformation.

In general, the structure and interactions in the XynB active site match the other family 11 xylanases very closely. Glu90 is hydrogen bonded to Tyr81 through one carboxyl O atom and to Gln139 through the other, and Glu180 is hydrogen bonded to Tyr92 through one carboxyl O atom and to Asn43 through the other. All these interactions involve residues that are fully conserved in other xylanases, except for Asn43, which varies to Asp in enzymes with a low pH optimum (Krengel & Dijkstra, 1996). The guanidinium group of Arg125 is close enough to maintain a low pK_a for Glu90, as proposed. The distance between the guanidinium and carboxyl groups varies in different xylanase structures, however, and in XynB is 5 Å in molecule *A* and 6.5 Å in molecule *B* (measured between the 90 C^{δ} and 125 C^{ζ} atoms). This flexibility implies that the role that Arg125 plays in activity may depend on the repositioning of this residue after the substrate has bound. Superposition of XynB on to the substrate complexes of *B. circulans* xylanase (Sidhu *et al.*, 1999) and *B. agaradhaerens* xylanase (Sabini *et al.*, 1999) appears to confirm this. The position the guanidinium group of Arg125 adopts in XynB would clash with the −1 subsite sugar and in both complexes the corresponding side chains are slightly displaced in order to hydrogen bond to the sugar.

Other active-site residues in XynB correspond remarkably closely in both position and orientation with the residues that contact the substrate sugar rings in the *B. circulans* and *B. agaradhaerens* substrate complexes (Sidhu *et al.*, 1999; Sabini *et al.*, 1999), suggesting that the substrate-binding site is essentially pre-formed. Taking

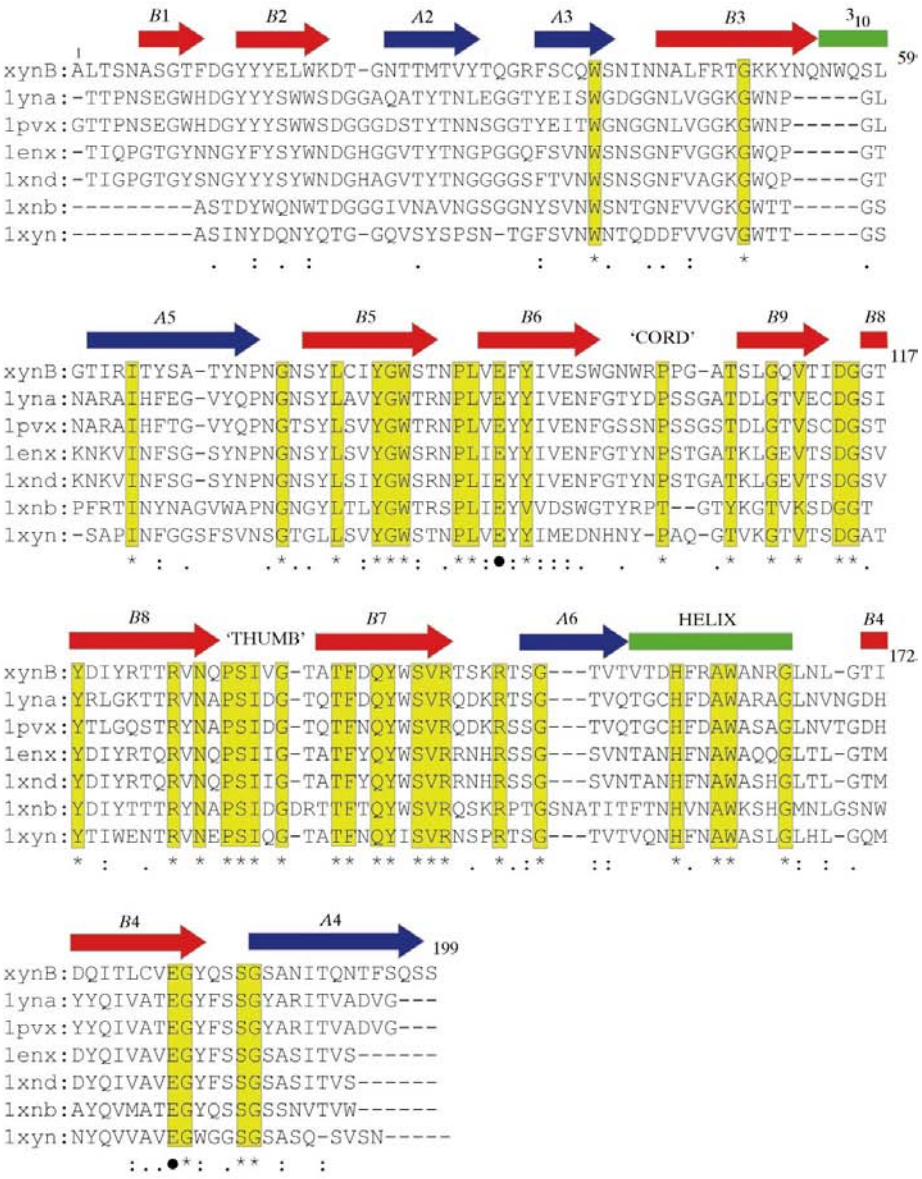

Figure 2
Multiple structure-based sequence alignment of the family 11 xylanases from *D. thermophilum* (XynB), *T. lanuginosus* (1yna), *P. varioti* (1pvx), *Tr. harzianum* (1xnd), *Tr. reesei* xylanases I (1xyn) and II (1enx) and *B. circulans* (1xnb). The location of secondary structures in the XynB structure are indicated. Residues that are identical in all seven enzymes are highlighted in yellow and marked (*). Residues that are strongly conserved (:) and those whose character is conserved (.) are also indicated. The catalytic Glu residues (90 and 180 in XynB) are indicated with a filled circle.

Table 3
Structural superpositions of family 11 xylanases.

The total number of residues in each protein is given on the diagonal. In the upper right half of the table are given the pairwise r.m.s. difference (Å) and the number of residues used in the superposition. In the lower left half is the pairwise percentage sequence identity.

Xylanase (PDB code)	XynB	*T. lanuginosus* (1yna)	*P. varioti* (1pvx)	*Tr. harzianum* (1xnd)	*B. circulans* (1xnb)	*Tr. reesei* I (1xyn)	*Tr. reesei* II (1enx)
XynB	199	1.01 170 C^{α}	1.06 175 C^{α}	1.01 171 C^{α}	1.06 164 C^{α}	1.07 158 C^{α}	1.18 171 C^{α}
T. lanuginosus	51	194	0.37 193 C^{α}	0.81 184 C^{α}	1.00 164 C^{α}	0.99 164 C^{α}	0.70 183 C^{α}
P. varioti	49	88	194	0.79 185 C	1.00 166 C	0.90 159 C	0.72 184 C
Tr. harzianum	56	60	60	190	0.84 166 C	0.81 157 C	0.69 188 C
B. circulans	48	53	53	56	185	1.14 166 C	1.13 167 C
T. reesei I	47	46	49	56	51	178	0.93 161 C
T. reesei II	56	62	61	94	54	57	190

the *B. agaradhaerens* complex (Sabini *et al.*, 1999) as a reference, these residues include in XynB Trp18, equivalent to Trp19, which stacks against the −2 subsite sugar ring in the complex; Glu16, Arg47 and Tyr81, equivalent to Glu17, Arg49 and Tyr85, which hydrogen bond to the −2 sugar hydroxyl groups in the complex; Leu45, equivalent to Leu47, which makes close van der Waals contacts with the sugar and may play a role in determining xylose sugar specificity; Trp83, Pro129 and Phe137, equivalent to Trp87, Pro133 and Phe141, which complete the definition of the binding site in the complex. Most of these residues are conserved in other xylanases, although Arg47 is usually a glycine and its role in hydrogen bonding to the substrate appears to be taken by a tyrosine (Tyr166 in *B. circulans*, which corresponds to Gln174 in XynB and Glu178 in *B. agaradhaerens*).

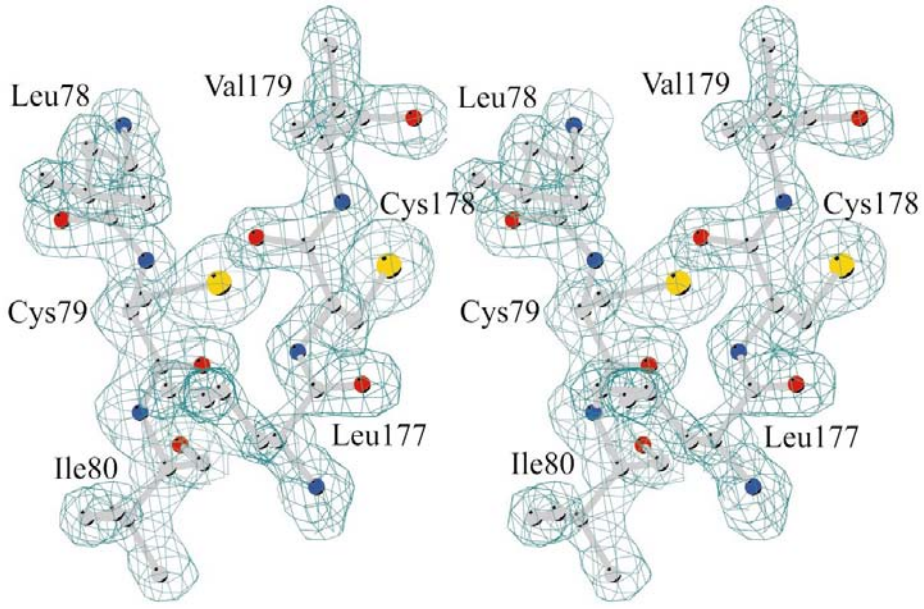

Figure 3
Stereo figure showing the electron density in the vicinity of the two cysteine residues Cys79 and Cys178; the side chain of Cys178 is oriented away from Cys79 such that no disulfide bond is formed. Electron density taken from the final $2F_o - F_c$ map contoured at a level of 1.5σ.

3.3. Cysteine residues

Several of the other family 11 xylanases contain disulfide bonds. In the thermostable xylanases from *P. varioti* (Kumar *et al.*, 2000) and *T. lanuginosus* (Gruber *et al.*, 1998), a disulfide bond 110–154 connects the C-terminal end of strand *B*9 to the N-terminus of the α-helix and this has been correlated with their enhanced stability. In *A. niger* xylanase I (Krengel & Dijkstra, 1996), a disulfide bond 92–111 joins the cord to strand *B*8. In contrast, in XynB there are three Cys residues, Cys36, Cys79 and Cys178, but no disulfide bond is formed. Curiously, Cys79 and Cys178, which are on adjacent β-strands, *B*5 and *B*4, in the active-site region, appear close enough to form a disulfide, but the electron density shows clearly that they do not (Fig. 3); instead, both sulfhydryl groups hydrogen bond to carbonyl O atoms, of residues 178 and 46, respectively. The distance between their C^{α} atoms is 5.2 Å, within the range found for other disulfide bonds (Thornton, 1981), and a change in the rotamer for Cys178 would bring the two S atoms within 2.5 Å. It may be the constraints of the β-structure that prevents disulfide formation, however, and Thornton's analysis (Thornton, 1981) shows that disulfide bonds occur relatively infrequently where both Cys residues have a β conformation. In the present case, disulfide-bond formation would require some main-chain adjustment to bring the S^{γ} atoms close enough in a favourable geometry and this could disrupt the mutual hydrogen bonding between the main-chain C═O and NH groups of Cys79 and Cys178. The third cysteine residue, Cys36, is buried and makes only non-bonded contacts with other residues; it is replaced by hydrophobic residues in other xylanases.

3.4. Solvent structure

Calculations of solvent accessibility show that associated with each molecule of XynB are seven water molecules that are effectively buried. One of these (W8 for molecule *A*, W3 for molecule *B*) is truly internal, being bound to the side chains of Asp114,

Table 4
Thermostability factors.

Xylanase	XynB	*T. lanuginosus*	*P. varioti*	*Tr. harzianum*	*B. circulans*	*Tr. reesei* I	*Tr. reesei* II
Polar surface (%)	83	73	73	73	78	75	73
Ion pairs	9	9	9	12	5	5	10
Hydrogen bonds (main-chain–main-chain)	122	118	114	119	116	113	120
Hydrogen bonds (main-chain–side-chain)	49	45	51	44	43	33	51
Hydrogen bonds (side-chain–side-chain)	44	38	38	40	39	27	36
Arg/Lys ratio	2.5	2.7	3.0	1.5	1.4	1.5	1.5
Asn + Gln residues	29	20	21	25	23	29	30
Pro residues	5	6	6	6	6	6	7
Gly residues	19	29	31	27	25	22	27
Cys residues	3	2	2	—	—	—	—
Aromatic residues	28	30	30	31	30	21	31

Tyr118 and His158 in the interface between the cord, the *B*8–*B*9 loop and the α-helix. This water molecule is conserved in all family 11 xylanases so far analysed and appears to make an important contribution to structure and stability (Kumar *et al.*, 2000).

None of the other buried waters are similarly conserved, although one (W6 molecule *A*, W9 molecule *B*) is present in all except the *Tr. reesei* xylanase I; this water molecule is in a crevice between the *A*5–*B*5 loop and the *B*6–cord connection, bound in XynB to Asn75 $O^{\delta1}$, 76 N, 100 N and 100 O. The other buried waters are more specific to XynB, mostly occupying crevices close to the surface and tying together the connecting loops between secondary-structure elements. One (W1 molecule *A*, W7 molecule *B*) ties the α-helix to the cord by hydrogen bonding to main-chain atoms of residues 99, 163 and 167; another (W29 molecule *A*, W25 molecule *B*) ties the *B*1−*B*2 loop to the *A*2−*A*3 loop by hydrogen bonding to main-chain atoms of residues 10, 29 and 32; a third (W54 molecule *A*, W66 molecule *B*) is buried between Trp56 from the *B*3–*A*5 loop and the *B*4–α-helix connection.

Two sulfate ions have been located in the crystal structure, both associated with the side chain of Arg165 in the active-site cleft of molecule *B*, but they do not seem likely to have any functional significance.

3.5. Thermostability

Various factors have been identified as contributing to enhanced protein thermostability (Vogt *et al.*, 1997; Wallon *et al.*, 1997; Knapp *et al.*, 1997). These include specific features such as additional disulfide bonds, metal binding, enhanced hydrogen bonding, increased numbers of ion pairs, oligomerization, shortened loops and shortened or protected N- and C-termini and an increased fraction of polar surface. Compositional factors such as more proline residues or fewer glycine residues, an increased Arg/Lys ratio or decreased numbers of residues prone to chemical change (Asn, Gln, Cys) may also contribute. The energy difference between mesophilic and thermophilic proteins is very small, however, and different factors evidently operate in different proteins.

In the family 11 xylanases an additional disulfide bridge, 110–154, has been associated with the enhanced thermostability of the enzymes from *T. lanuginosus* (Gruber *et al.*, 1998) and *P. varioti* (Kumar *et al.*, 2000). This disulfide links strand *B*9 to the α-helix in the hinge region at the back of the molecule (Fig. 4) and it is suggested that enhanced interactions in this region, including also some specific hydrogen bonds, are important for stability. Apart from this, however, there are no obvious stabilizing factors.

XynB is the most thermostable xylanase yet to be structurally characterized, having a temperature optimum of 348 K (Morris *et al.*, 1998) compared with 343 K for *T.*

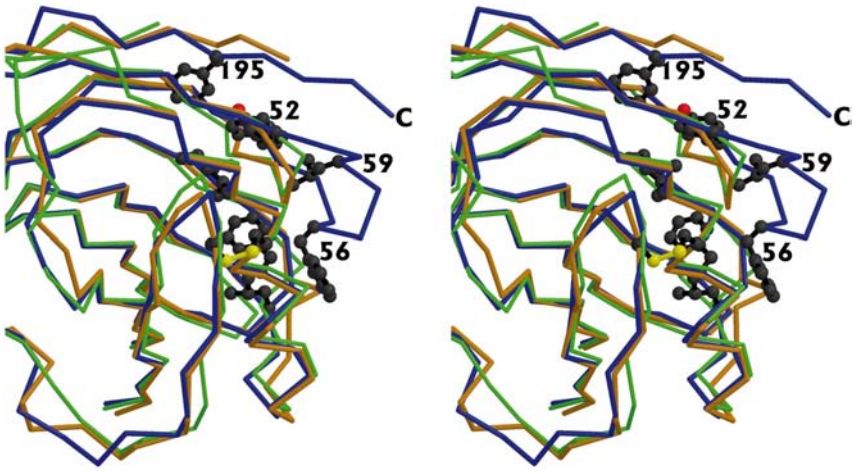

Figure 4
Stereo diagram showing the structure in the vicinity of the insertion in the *B*3–*A*5 loop in XynB from *D. thermophilum*. In XynB (blue) this loop, including the 3_{10}-helix 55–59, folds in against the body of the molecule such that an extended hydrophobic core, including Tyr52, Leu59 and Ile62, is covered by the partly buried Trp56 from the centre of the loop. Phe195 from the extended C-terminus (top of picture) can also be seen, inserting into the core and tying down the C-terminal strand (C). In contrast, in the *P. varioti* xylanase (gold) and *B. circulans* xylanase (green) xylanases, the loop is five residues shorter and includes several exposed hydrophobic residues (not shown). The disulfide bond 110–154, which links the helix to strand B9 in the *P. varioti* xylanase and has been associated with enhanced thermostability, is also shown (yellow spheres).

lanuginosus xylanase (Gruber *et al.*, 1998). We can identify several factors that may be responsible for its increased thermostability. Firstly, the percentage of solvent-accessible surface that is polar, calculated using *GRASP* (Nicholls *et al.*, 1991), is significantly higher than in the other xylanases; for XynB the surface is 83% polar, compared with 78% for *B. circulans* xylanase and 73–75% for all the others (Table 4). An increased proportion of polar surface (and corresponding decrease in exposed hydrophobic surface) correlates strongly with increased thermostability (Vogt *et al.*, 1997) and the figure of 83% for XynB is similar to that found for highly thermostable enzymes of the subtilisin family (Smith *et al.*, 1999).

Secondly, the altered *B*3–*A*5 connecting loop and the associated extended C-terminus, shown in Fig. 4, seem likely to provide a local contribution to the increased thermostability. In contrast to the other family 11 xylanases, in XynB this loop is five residues longer and contains a short 3_{10}-helix that tucks back in to pack against the α-helix and side chains from strands *A*5 and *B*4. This produces new hydrophobic interactions; Leu59 and Ile62 from the *B*3–*A*5 loop are buried, packing against Tyr52, Val155, Phe159, Ile175 and Phe195, and Trp56 at the centre of the loop is partially buried and contacts Phe159, Leu169 and Ile177, effectively closing off this extended hydrophobic core. Phe195 from the extended C-terminus also tucks into this hydrophobic core, in a position occupied by the side chain of a Trp residue in other xylanases (Trp42 in *B. circulans* xylanase). This may help to tie down the C-terminal strand. In contrast to XynB, the shorter *B*3–*A*5 loop in other xylanases leaves several hydrophobic side chains exposed; for example, Pro47 and Phe48 in *B. circulans* xylanase and Pro53 and Leu55 in *P. varioti* xylanase.

The larger *B*3–*A*5 loop also results in enhanced hydrogen bonding, both through the formation of the 3_{10}-helix 55–59 and because of increased β-sheet hydrogen bonding. Residues 61–64 from the *B*3–*A*5 loop hydrogen bond with residues 196–198 of the extended C-terminal strand *A*4 on one side and with the slightly extended strand *A*6 on the other. The result is that the *B*3–*A*5 loop in XynB makes 34 hydrogen bonds, compared with 20 for the corresponding loop in *B. circulans* xylanase. Extended hydrogen bonding may also explain the even higher thermostability of another variant of XynB. This variant, XynB6, has an N-terminal extension of six residues (Morris *et al.*, 1998), which may extend the β-sheet structure of the molecule.

Local structural changes can enhance thermostability, particularly if the region is a potential 'hot-spot' that may nucleate unfolding or increase flexibility. In the family 11 xylanases, the *B*3–*A*5 loop, implicated here in the increased thermostability of XynB, and the disulfide bridge 110–154, implicated in the thermostability of the *T. lanuginosus* and *P. varioti* enzymes, are both close to the hinge region at the back of the molecule. This seems likely to be a critical 'hot-spot' in unfolding. Indeed, engineered disufide bridges in this region of the *B. circulans* xylanase have been shown to enhance thermostability by up to 15° (Wakarchuk, Sung *et al.*, 1994).

In most other respects, XynB resembles the other family 11 xylanases. The number of ion pairs defined with a distance cutoff of 4.0 Å (Barlow & Thornton, 1983) is similar in all the proteins. The numbers of hydrogen bonds defined using the criteria of Baker & Hubbard (1984) are also similar, although XynB does have slightly more hydrogen bonds in total than any of the other proteins, in addition to the local increase noted above. XynB also has significantly fewer glycine residues than the others (19 compared with 25–30 in most of the others), but similar numbers of Pro, Asn, Gln and aromatic residues and no significant difference in the ratio of Arg/Lys. A curious feature is that XynB alone in this family of xylanases has three unpaired Cys residues. This might be expected to decrease thermostability because of their susceptibility to oxidation, but we assume that their burial in XynB militates against this.

We gratefully acknowledge Dr Clyde Smith and staff of the Stanford Synchrotron Radiation Laboratory for help with data collection, and Lisa Harris for providing the purified protein. This work was supported by the Marsden Fund of New Zealand (grant to ENB), the Wellcome Trust (major equipment grant to ENB) and the Australian Research Council (grant to PLB).

References

Asselt, E. J. van, Perrakis, A., Kalk, K. H., Lamzin, V. S. & Dijkstra, B. W. (1998). *Acta Cryst.* D**54**, 58–73.
Baker, E. N. & Hubbard, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.
Barlow, D. J. & Thornton, J. M. (1983). *J. Mol. Biol.* **168**, 867–885.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
Brunger, 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.
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.
Cowtan, K. & Main, P. (1998). *Acta Cryst.* D**54**, 487–493.
Davies, G. & Henrissat, B. (1995). *Structure*, **3**, 853–859.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Gebler, J., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Berguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Warren, R. A. J. & Withers, S. G. (1992). *J. Biol. Chem.* **267**, 12559–12561.
Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W. & Kratky, C. (1998). *Biochemistry*, **37**, 13475–13485.
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.
Henrissat, B. (1991). *Biochem. J.* **280**, 309–316.
Henrissat, B. & Bairoch, A. (1993). *Biochem. J.* **293**, 781–788.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Keitel, T., Simon, O., Borriss, R. & Heinemann, U. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 5287–5291.
Knapp, S., de Vos, W. M., Rice, D. & Ladenstein, R. (1997). *J. Mol. Biol.* **267**, 916–932.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Krengel, U. & Dijkstra, B. W. (1996). *J. Mol. Biol.* **263**, 70–78.
Kumar, P. R., Eswaramoorthy, S., Vithayathil, P. J. & Viswamitra, M. A. (2000). *J. Mol. Biol.* **295**, 581–593.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lu, G. (1996). *Protein Data Bank Quart. Newslett.* **78**, 10–11.
Lu, G. (1998). *MAPS: Multiple Alignment of Protein Structures.* http://bioinfo1.mbfys.lu.se/TOP/maps.html.
McCarter, J. D. & Withers, S. G. (1994). *Curr. Opin. Struct. Biol.* **4**, 885–892.
McDonald, I. K. & Thornton, J. M. (1994). *Protein Eng.* **8**, 217–224.
McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Körner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W. & Withers, S. G. (1996). *Biochemistry*, **35**, 9958–9966.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Morris, D. D., Gibbs, M. D., Chin, C. W., Koh, M. H., Wong, K. K., Allison, R. W., Nelson, P. J. & Bergquist, P. L. (1998). *Appl. Environ. Microbiol.* **64**, 1759–1765.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.
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. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.
Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jorgensen, P. L., Schülein, 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.
Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M. & Baker, E. N. (1999). *J. Mol. Biol.* **294**, 1027–1040.
Thornton, J. M. (1981). *J. Mol. Biol.* **151**, 261–287.
Törrönen, A., Harkki, A. & Rouvinen, J. (1994). *EMBO J.* **13**, 2493–2501.
Törrönen, A. & Rouvinen, J. (1995). *Biochemistry*, **34**, 847–856.
Viikari, L., Kantelinen, A., Sundquist, J. & Linko, M. (1994). *FEMS Microbiol. Rev.* **13**, 335–350.
Vogt, G., Woell, S. & Argos, P. (1997). *J. Mol. Biol.* **269**, 631–643.
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.
Wallon, G., Kryger, G., Lovett, S. T., Oshima, T., Ringe, D. & Petsko, G. A. (1997). *J. Mol. Biol.* **266**, 1016–1031.