Acta Cryst. (1996). D**52**, 393–401

Refined Crystal Structure of the Catalytic Domain of Xylanase A from *Pseudomonas fluorescens* at 1.8 Å Resolution

GILLIAN W. HARRIS, JOHN A. JENKINS, IAN CONNERTON AND RICHARD W. PICKERSGILL

Department of Protein Engineering, Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road, Reading RG6 6BZ, England

(*Received 21 July 1995; accepted 3 October 1995*)

Abstract

The three-dimensional structure of native xylanase A from *Pseudomonas fluorescens* subspecies *cellulosa* has been refined at 1.8 Å resolution. The space group is $P2_12_12_1$ with four molecules in the asymmetric unit. The final model has an R factor of 0.166 for 103 749 reflections with the four molecules refined independently. The tertiary structure consists of an eightfold β/α-barrel, the so-called TIM-barrel fold. The active site is in an open cleft at the carboxy-terminal end of the β/α-barrel, and the active-site residues are a pair of glutamates, Glu127 on strand 4 and Glu246 on strand 7. Both these catalytic glutamate residues are found on β-bulges. An atypically long loop after strand 7 is stabilized by calcium. Unusual features include a non-proline *cis*-peptide residue Ala80 which is found on a β-bulge at the end of β-strand 3. The three β-bulge type distortions occurring on β-strands 3, 4 and 7 are functionally significant as they serve to orient important active-site residues. The active-site residues are further held in place by an extensive hydrogen-bonding network of active-site residues in the catalytic site of xylanase A. A chain of well ordered water molecules occupies the substrate-binding cleft, some or all of which are expelled on binding of the substrate.

1. Introduction

Endo-1,4-β-D-xylanase (E.C. 3.2.1.8) catalyzes the random internal cleavage of the β-1,4-glycosidic bonds of xylan, the second major plant cell-wall polysaccharide. Cellulases and xylanases have been classified into several families on the basis of sequence homology of the catalytic binding domains (Gilkes, Henrissat, Kilburn, Miller & Warren, 1991). The catalytic domains of known xylanases can be assigned to either family F or G, which correspond to families 10 and 11 of the Henrissat (1991) classification of glycosyl hydrolases. The three-dimensional structures of low molcular weight xylanases (family G; $M_r \simeq 20$ kDa) from *Bacillus pumilus* (Arase *et al.*, 1993; Dauter, 1993), *Bacillus circulans* and *Trichoderma harzianum* (Campbell *et al.*, 1993) and *Trichoderma reesei* (Törronen, Harkki & Rouvinen, 1994; Törronen & Rouvinen, 1995) have been reported. These homologous enzymes are single domain β-sheet proteins. On the basis of crystal structure results and mutational studies, a lysozyme-type mechanism has been proposed for family G xylanases (Wakarchuk, Campbell, Sung, Davoodi & Yaguchi, 1994).

The higher molecular weight xylanases ($M_r \simeq 32$–39 kDa) belong to family F (glycosyl hydrolase family 10). Family F xylanases consist of a cellulose-binding domain and a catalytic domain connected by a flexible linker region. Several structures of the catalytic binding domain of family F xylanases have been reported in the past year: β-1,4-glycanase Cex from *Cellulomonas fimi* (White, Withers, Gilkes & Rose, 1994), *Streptomyces lividams* xylanase A (Derewenda *et al.*, 1994), xylanase Z from *Clostridium thermocellum* (Dominguez *et al.*, 1995) and xylanase A (refinement of active site E246C mutant at 2.5 Å and preliminary refinement of native at 3.0 Å resolution) from *Pseudomonas fluorescens* (Harris *et al.*, 1994). These structures reveal that the catalytic domain of the F-family xylanase is an eightfold β/α-barrel.

Here we describe and discuss the refined X-ray crystal structure at 1.8 Å resolution of the catalytic core of the family F xylanase A ($M_r = 39$ kDa) from *Pseudomonas fluorescens* subspecies *cellulosa*. The active-site residues are a pair of glutamates, Glu127 on strand 4 and Glu246 on strand 7 of the eightfold β/α-barrel (Fig. 1). These glutamates are conserved across homologous family F xylanases. Family F xylanases are themselves a family member of a superfamily of glycohydrolase enzymes with eightfold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands 4 and 7 (Jenkins, Lo Leggio, Harris & Pickersgill, 1995). The 4/7 superfamily includes families 1 (β-glucosidases), 2 (β-galactosidases), 5 (family A cellulases), 10 (family F xylanases) and 17 (β-glucanases).

2. Methods

2.1. *Data collection*

The expression and crystallization of the catalytic domain of xylanase A from *Pseudomonas fluorescens* subsp. *cellulosa* has been reported elsewhere (Pickersgill *et al.*, 1993). The xylanase A crystallized in the

Table 1. *Intensity data statistics: completeness and multiplicity versus resolution for xylanase A*

Range	d_{min} (Å)	No. measured	No. unique reflections	% Possible	multiplicity	$R(I)$	I/σ
1	6.92	26829	2253	89.1	11.9	0.051	10.5
2	4.90	57315	4273	97.8	13.4	0.054	10.8
3	4.00	69401	5378	96.7	12.9	0.056	10.5
4	3.46	74980	6032	92.6	12.4	0.061	10.1
5	3.10	75836	6399	87.1	11.9	0.071	9.0
6	2.83	69972	6942	85.8	10.1	0.087	7.7
7	2.62	53770	7437	84.9	7.2	0.101	6.9
8	2.45	41921	7776	82.8	5.4	0.114	6.1
9	2.31	36225	8051	80.6	4.5	0.131	5.5
10	2.19	34925	8231	78.1	4.2	0.151	4.8
11	2.09	34647	8415	76.0	4.1	0.181	4.0
12	2.00	34172	8579	74.0	4.0	0.222	3.3
13	1.92	33378	8622	71.5	3.9	0.280	2.6
14	1.85	32705	8816	70.3	3.7	0.373	2.0
15	1.79	27549	7963	61.5	3.5	0.479	1.5
Overall		703625	105167	78.8	6.7	0.077	7.6

orthorhombic space group $P2_12_12_1$ with $a = 96.11$, $b = 97.32$, $c = 151.03$ Å and with four molecules in the asymmetric unit. Data to 1.8 Å resolution [$R_{sym}(I)$ of 0.077, 78.8% complete] were collected using the Weissenberg camera on beamline BL6A at the Photon Factory, Japan (Sakabe, 1991). The data were processed using the program *DENZO* (Otwinowski, 1993). Intensity data statistics are given in Table 1.

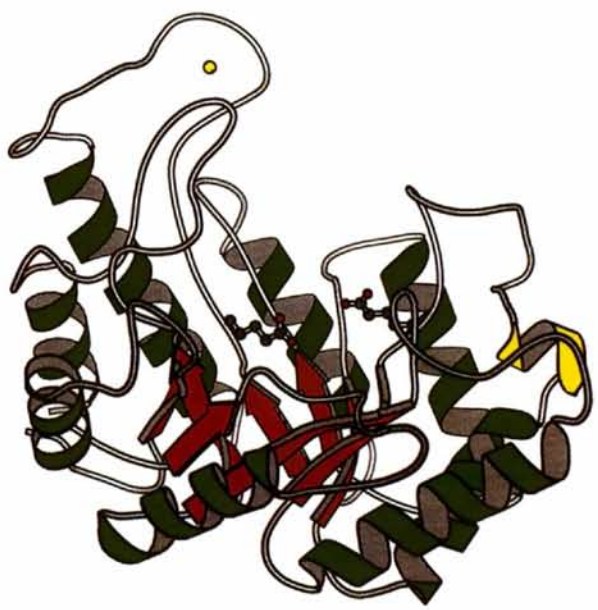


Fig. 1. The eightfold β/α-barrel of xylanase A viewed perpendicular to the barrel axis showing the acid base Glu127 on strand 4 and the nucleophile Glu246 on strand 7. The Ca atom (yellow dot) stabilizes the long loop after strand 7. This figure was prepared using *MOLSCRIPT* (Kraulis, 1991). The β-strands are shown as arrows in red and the eight α-helicies are shown in green; the additional α-helix ($\alpha 4a$) is shown in yellow. The glutamates are 5.5 Å apart consistent with a retention of the configuration during catalysis.

Table 2. *Refinement and stereochemical statistics of xylanase A at 1.8 Å*

(*a*) Refinement statistics

Resolution range (Å)	10.0–1.8
No. of reflections	103749
No. of protein atoms	10800
No. of solvent sites	1108
No. of bound atoms (Ca)	4
Residual, R*	0.166
Weighted residual, wR†	0.176
Correlation coefficient C‡	0.959
No. positional parameters	35736
No. thermal parameters	11912
No. overall scaling parameters	2
No. occupancies refined (Ca)	4
Total no. parameters refined	47654
Total no. restraints	27968
Total no. of data	131717
Mean B value for protein atoms (Å^2)	16.6

(*b*) Stereochemical statistics

Distances D (Å)	Number	R.m.s. deviation (Å) from ideal values
$D < 2.12$	10712	0.008
$2.12 < D < 2.62$	14412	0.027
$D > 2.62$	156	0.041
Planes		
Type 1 (peptide)	1380	0.012
Type 2 (other)	588	0.005
Chiral centres§	720	0.011

* $R = \sum |F_o| - G|F_c| / \sum |F_o|$. † $wR = [|\sum w(|F_o| - G|F_c|)^2 / \sum w|F_o|^2|]^{1/2}$
‡ $C = n\sum(|F_o|G|F_c|) - \sum |F_o| \sum G|F_c| / \{[n\sum |F_o|^2 - (\sum |F_o|)^2] \times [n\sum(G|F_c|)^2 - (\sum G|F_c|)^2]\}^{1/2}$, where n is the number of amplitudes used, G is the scale applied to F_c, and constant weight, $w = 0.1 \times 10^{-3}$, given to all reflections. § Chiral restrains were applied as distance restraints along the edges of chiral tetrahedra with $d' > 2.12$ Å.

2.2. *Structure solution*

The structure of native xylanase A was determined by molecular replacement using the program *AMoRe* (Navaza, 1994) using the 2.5 Å structure of the E246C mutant xylanase A as the search model and using data to 3 Å (Harris *et al.*, 1994). The rotation-function peaks corresponding to the positions of the four molecules in the asymmetric unit were at least 40% higher than the next highest peak. The correct translation solutions were selected on the basis of the highest correlation coefficient: 0.186 for molecule 1, 0.364 for molecule 2 with molecule 1 fixed, 0.542 for molecule 3 with molecules 1 and 2 fixed, rising to 0.729 molecule 4 with molecules 1, 2 and 3 fixed. Rigid-body refinement of the four molecules in the asymmetric unit using the program *RESTRAIN* (Driessen *et al.*, 1989) gave an R factor of 0.28 using data to 3.0 Å.

2.3. *Crystallographic refinement*

The structure was refined using the program *RESTRAIN*. Initially atomic positions of the four molecules were refined applying strict non-crystallo-

graphic symmetry (R factor of 0.30 at 2.0 Å). The resolution was extended to 1.8 Å and individual isotropic atomic temperature factors were also refined (R factor of 0.26). A *SIGMAA*-weighted (Read, 1986) $2F_o - F_c$ map at 1.8 Å resolution was calculated (Collaborative Computational Project, Number 4, 1994) and manual rebuilding using *O* (Jones, Zou, Cowan & Kjeldgaard, 1991) on an Evans and Sutherland PS390 graphics system, followed by refinement using *RESTRAIN* reduced the R factor to 0.23. Further refinement failed to reduce the R factor. The four molecules were then refined independently and the R factor dropped to 0.19. Further cycles of rebuilding molecule one to *SIGMAA*-weighted maps, regenerating the other three molecules by application of the non-crystallographic symmetry matrices, followed by positional and individual isotropic atomic temperature-factor refinement of the four molecules independently using *RESTRAIN* reduced the R factor to 0.166 for a model with one Ca atom and 277 water molecules per protein molecule (345 residues). The last two C-terminal residues (346 and 347) could not be seen in the electron density and were not included in the model. Refinement statistics are given in Table 2(*a*). Stereochemical statistics for the model are given in Table 2(*b*) and a Ramachandran plot is shown in Fig. 2 and plot statistics are in Table 3; all the residues fall within the favoured (88.9%) or additionally allowed (11.1%) regions (*PROCHECK*, Laskowski, MacArthur, Moss & Thornton, 1993).

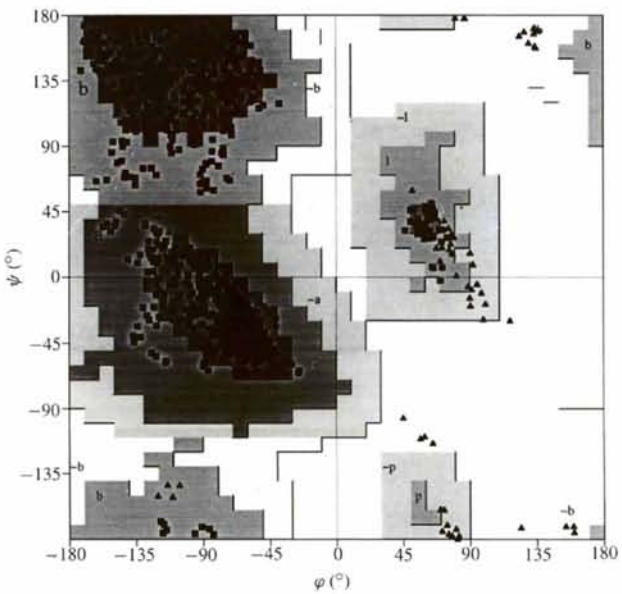


Fig. 2. Ramachandran plot for xylanase A at 1.8 Å produced by *PROCHECK* (Laskowski *et al.*, 1993). The 96 glycine residues are shown as triangles.

Table 3. *Ramachandran plot statistics*

Based on an analysis of 118 structures of at least 2.0 Å and R factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Residues in most favoured regions [A,B,L] (No., %)	1078	88.9
Residues in additional alowed regions [a,b,l,p] (No., %)	134	11.1
Residues in generously allowed regions [~a, ~b, ~l, ~p] (No., %)	0	0.0
Residues in disallowed regions (No., %)	0	0.0
No. of non-glycine and non-proline residues (No.,%)	1212	100.0
No. of end-residues (excl. Gly and Pro)	4	
No. of glycine residues (shown as triangles)	96	
No. of proline residues	68	
Total No. of residues	1380	

2.4. *Comparison to β-1,4-glycanase Cex*

Coordinates of a homologous family F structure, the β-1,4-glycanase Cex (Cex) from *Cellulomonas fimi* (White *et al.*, 1994) are available from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) (PDB code 2EXO). The quality of the refined Cex structure (R factor of 0.217 at 1.8 Å resolution) is comparable to xylanase A. These Cex coordinates were structurally aligned with the xylanase A (molecule 1) coordinates using the program *MNYFIT* 5.0 (Sutcliffe, Haneef, Carney & Blundell, 1987), which gave an r.m.s. deviation of 1.12 Å for 238 equivalent Cα atoms. The superimposed molecules were displayed using *Quanta* 4.0 (Molecular

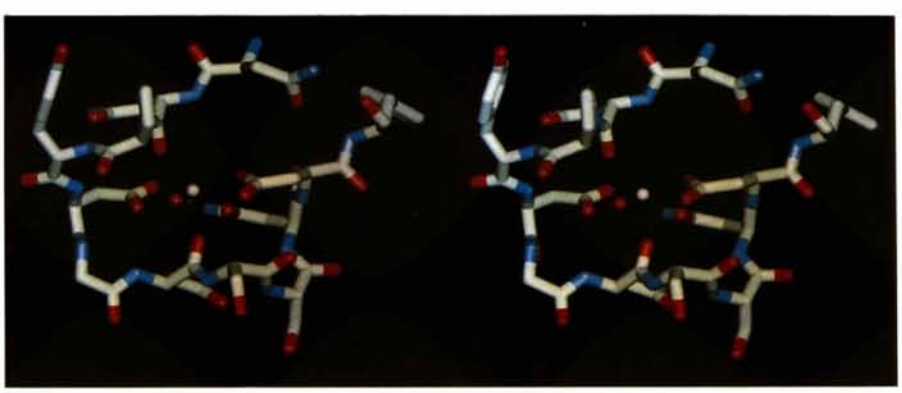


Fig. 3. Stereoview of the calcium-binding site of xylanase A. Calcium is shown as the white sphere and water OW349 as the red sphere.

Simulations Inc., 1994) run on a Silicon Graphics Indigo^2 Elan.

3. Results and discussion

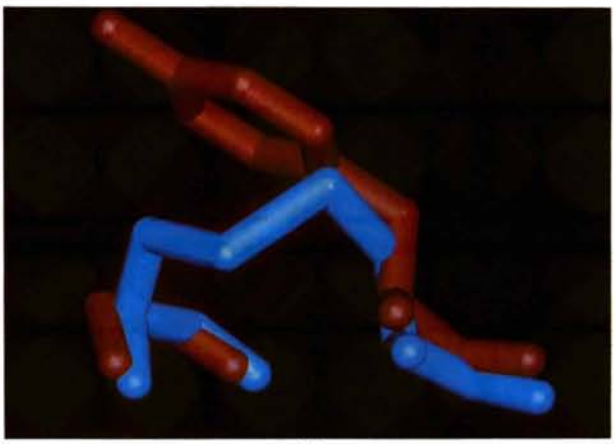
The protein has an eightfold β/α-barrel fold (Fig. 1), the so-called TIM-barrel structure, after triose phosphate isomerase, the first known example of this architecture (Banner *et al.*, 1975). This is a common fold for an enzyme, about one in ten enzymes having this architecture (Farber, 1993) and is found in homologous family F xylanase structures (White *et al.*, 1994; Derewenda *et al.*, 1994; Dominguez *et al.*, 1995). Cellobiohydrolase, a family B cellulase, has a related β/α-barrel fold (Rouvinen, Bergfors, Teeri, Knowles & Jones, 1990). Eight-stranded β/α-barrels have previously been classified on the basis of their shape (Farber & Petsko, 1990). Although xylanase A has the archetypal $(\beta/\alpha)_8$ fold, the β/α-barrel differs from the elliptical barrel structure seen in triose phosphate isomerase which has its major axis between strands 3 and 7. The shape of the barrel in xylanase A is elliptical in cross section with the major axis running between β-strands 1 and 5.

The β-sheets are residues 10–14 (β1), 39–42 (β2), 75–80 (β3), 121–126 (β4), 175–180 (β5), 209–212 (β6*a*), and 215–217 (β6*b*), 241–250 (β7) and 302–304 (β8), and the α-helices are 27–36 (α1), 60–73 (α2), 100–118 (α3), 149–154 (α4*a*) and 159–170 (α4*b*), 188–203 (α5), 223–234 (α6), 274–293 (α7) and 336–344 (α8). Xylanase A has one additional helix (α4*a*) and one additional β-sheet (β6*b*) as well as an atypically long loop (not present in homologous xylanase sequences) after strand 7 that is stabilized by calcium.

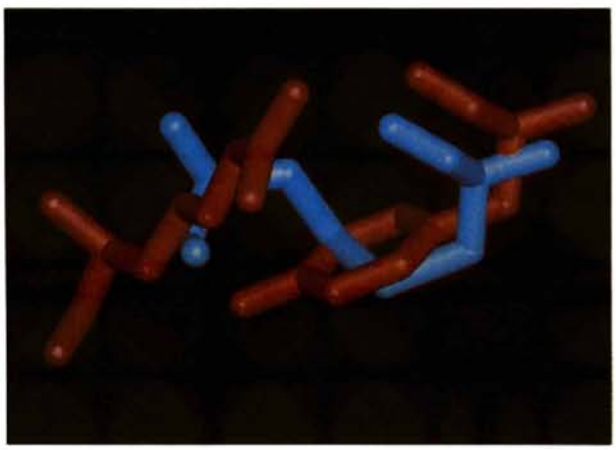
The active site is in an open cleft at the carboxy-terminal end of the β/α-barrel. The active-site residues are a pair of glutamates, the nucleophile Glu246 on strand 7 and the acid-base Glu127 at the end of strand 4 (Fig. 1). Both these residues are highly conserved in homologous xylanases. This work confirms the identity of these glutamates as active-site residues as shown by kinetic and mutagenesis experiments (MacLeod, Lindhorst, Withers & Warren, 1994; Tull & Withers, 1994; Lee, Lowe, Henrissat & Zeikus, 1993). The catalytic glutamates Glu246 and Glu127 are separated by a distance of 5.5 Å, consistent with a retention of conformation reaction mechanism.

The calcium-binding site is located on the long loop after β-strand 7 which bears the nucleophile Glu246. Ligands are contributed by Asp256 (OD1 and OD2), Asn261 (OD1), Asp262 (OD1), carbonyl O atoms of 253 and 258 and a water molecule (OW349). (Fig. 3). The disulfide bond between Cys269 and Cys273 is also situated on part of this loop, immediately before α-helix α7. This loop constitutes an insertion which does not occur in the sequences of homologous xylanases. The disulfide bridge forms a five-membered loop (Cys269-Ala-Val-Ser-Cys273) at the end of the long loop, immediately before α-helix α7. This arrangement places Val271 in an unusually exposed position at the apex of the five-membered loop formed by the disulfide bridge.

The mean isotropic B value for the 10800 protein atoms is 16.6 Å^2. The core of the xylanase A molecule is well ordered and the atoms with higher B values are in general those at the ends of polar residues which are exposed on the surface of the molecule. The regions of the structure which have the highest B values are the loops after β-strands β2 (residues 53–56), β3 (residues 95–102), β4 (residues 132–139) and β8 (residues 316–319). The two latter regions are found in the interface between the two xylanase A molecules which are related by a non-crystallographic approximate twofold axis (see below). The atoms of the long loop after β-strand β7 do not have high B values despite the exposed nature of this



(*a*)



(*b*)

Fig. 4. Superposition of xylanase A and Cex showing the two disulfide bonds in Cex (blue) which are not present in xylanase A (red). (*a*) The disulfide bond between Cys167 and Cys199 in Cex corresponds to residues Tyr177 and Gly209 in xylanase A. (*b*) The disulfide bond between Cys261 and Cys267 in Cex corresponds to residues Tyr290 and Arg299 in xylanase A. The internal gap in xylanase A which corresponds to the disulfide bridge in Cex is filled by the side chain of a tyrosine residue (*a*) Tyr177 and (*b*) Tyr290.

loop as the loop is stabilized by both the bound calcium and the disulfide bridge.

The single residue most notable for its high B values is Trp313. The end of this residue would appear to be flapping, as the atomic B values increase from low values (8–15 Å^2) for the CD1 and NE1 atoms to much higher values (25–77 Å^2) for the CE3, CZ2, CZ3 and CH2 atoms at the end of the ring furthest from the CG atom. This exposed tryptophan residue is partially anchored in position by a hydrogen bond (2.84 Å) between its NE1 atom and OD2 of Asp248. Asp248 is on β-strand β7, near the active-site glutamate Glu246. Both Asp248 and Trp313 are highly conserved across family F xylanases and substrate-binding studies of the inactive xylanase A E246C mutant have implicated Trp313 in binding of the xylopentaose substrate at binding site E (Harris *et al.*, 1994). The mobility of this tryptophan residue may be important for the binding of large xylan substrates in the substrate-binding site.

3.1. *Comparison of native xylanase A with β-1,4-glycanase Cex*

It is of interest to compare the structure of xylanase A to that of a homologous family F xylanase/glucanase, β-1,4-glycanase Cex (Cex) from *Cellulomonas fimi* (White *et al.*, 1994) in order to see which structural features are conserved, as these are likely to be of functional significance. Overall, the structure of Cex superimposes well with that of xylanase A, and the greatest differences in the main-chain traces are found, as expected, in the loop regions, especially the long loops at the C-terminal end of the β-strands. The loop after β-strand 4 is longer by nine residues in xylanase A, the conformation is completely different, and the additional α-helix α4*a* which follows, although also present in Cex, superimposes less well. Both xylanase A and Cex have the additional β-sheet β6*b* immediately following β-sheet 6 (β6*a*). The main difference is the loop after β-strand 7, which is substantially longer in xylanase A (24 residues, as compared to 4 in Cex) and incorporates the only

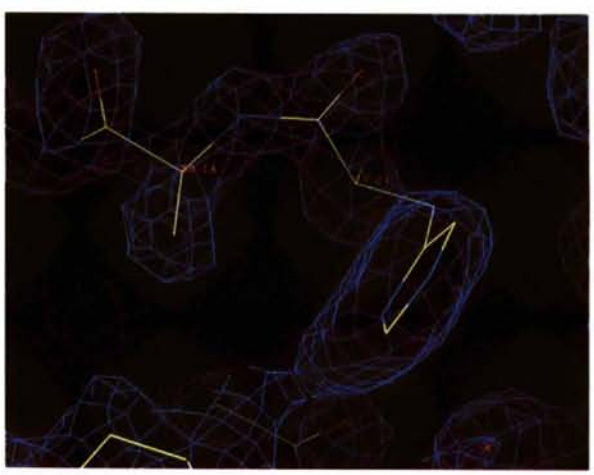


Fig. 5. The final model for the non-prolyl *cis*-peptide bond between the carbonyl of His79 and the amide of Ala80 together with a simulated-annealing omit map calculated using *X-PLOR* (Brünger, 1992). The atoms within a 6.0 Å sphere of Ala80 were omitted and the atoms within a 3.0 Å shell were restrained. Structure factors from *X-PLOR* were used to calculate the $2F_o - F_c$ map at 1.8 Å resolution which is contoured at 1σ.

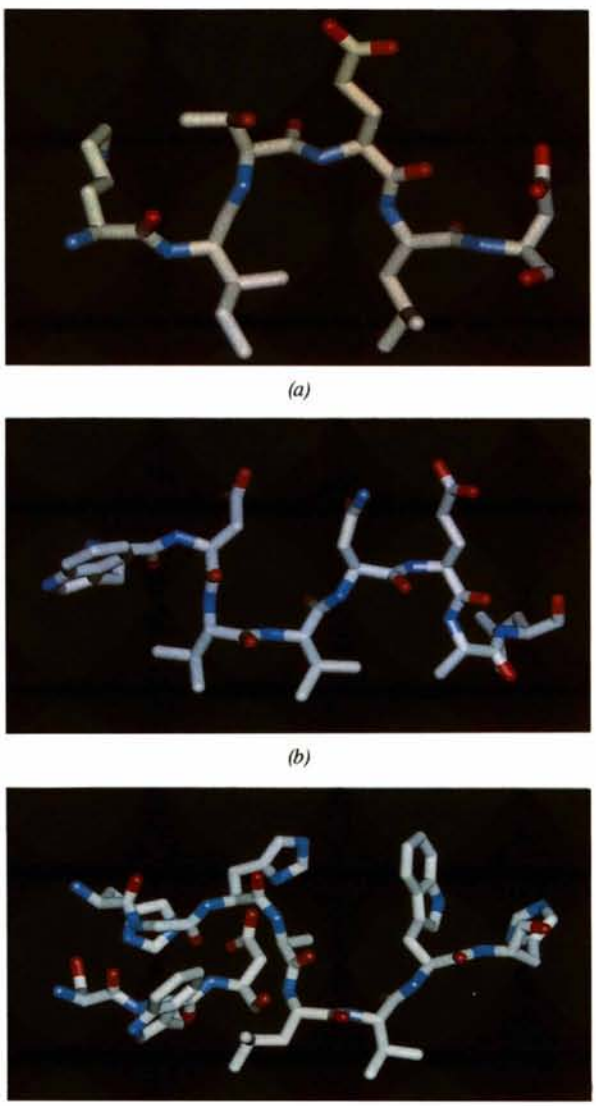


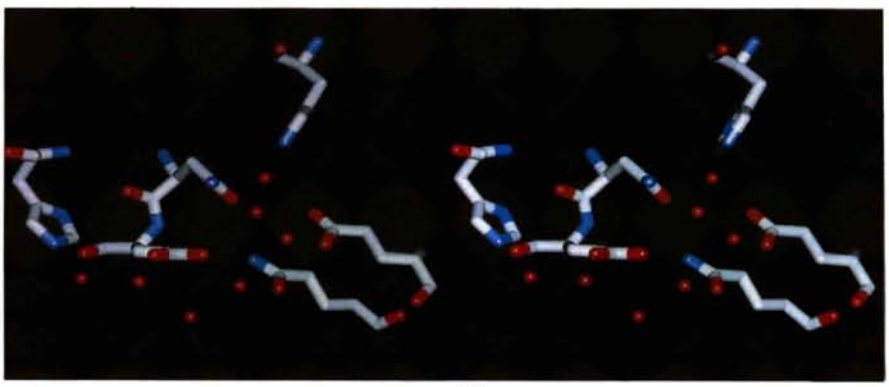
Fig. 6. The β-bulge features in xylanase A. (*a*) β-strand 6: single β-bulge involving residues Thr245 and nucleophile Glu246. (*b*) β-strand 4: double β-bulge crankshaft feature involving the residues Val124, Val125, Asn126 and acid–base catalyst Glu127. (*c*) β-strand 3: β-bulge (His79 and *cis*-Ala80) followed by a β-turn involving residues Ala80 Leu81, Val82 and Trp83. The space under Gly78 is filled by the side chain of Trp122 on the adjacent β-strand β4. The neighbouring residue Asp123 makes hydrogen bonds with both His77 and His79.

disulfide bond in xylanase A, as well as the calcium-binding site, which may serve to stabilize this long loop. The 32-residue loop after β-strand 8 is also substantially different. In Cex, this loop contains additional secondary structure, an extra α-helix ($\alpha 8$), followed by an extra β-strand ($\beta 9$). In xylanase A, the loop adopts a different conformation and α-helix $\alpha 8$ maps to the last α-helix $\alpha 9$ in Cex. In addition, Cex has short sections of α-helix, $\alpha 2a$ and $\alpha 3a$, on the loops before α-helix $\alpha 2(b)$ and $\alpha 3(b)$, respectively, and a short α-helix ($\alpha 0$) at the N-terminus, before β-strand 1, which are not formed in xylanase A.

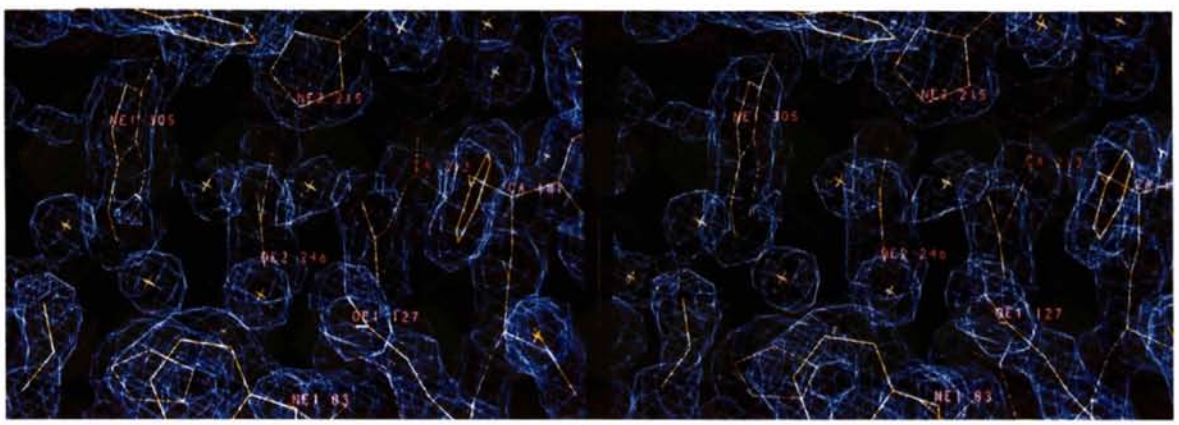
Cex contains two disulfide bonds which are not present in xylanase A. The disulfide bond between Cys167 and Cys199 links β-strand $\beta 5$ to the start of β-strand $\beta 6a$. The second disulfide bond is between Cys261 (in α-helix $\alpha 7$) and Cys267 at the start of β-helix $\beta 8$. This loop between α-helix $\alpha 7$ and β-strand $\beta 8$ is four residues longer in xylanase A. In both instances, the internal gap in xylanase A which corresponds to the disulfide bridge in Cex is filled by the side chain of a tyrosine residue in the equivalent position to the first cysteine of the disulfide pair, Tyr177 and Tyr290, respectively (Figs. 4*a* and 4*b*, respectively).

Xylanase A contains two *cis*-peptide residues, Ala80 and Pro221. *Cis*-Pro221 is on the short (five-residue) loop before α-helix $\alpha 6$, following on from the additional β-sheet $\beta 6b$. Although there is a proline residue (Pro212) in Cex in a similar position, this is not a *cis*-proline and the loops (two residues longer in Cex) are different in this region. The preceding Tyr220 side chain is held in place by the OH forming a hydrogen bond to Glu185. These residues are not correspondingly present in Cex. The *cis*-conformation arises from the need to avoid a steric clash with the tyrosine ring. The *cis*-Pro221 is stabilized by Arg282, Arg282 NH2 is within hydrogen-bonding distance of the carbonyl O atoms of Tyr220 and Pro221. Cex contains a *cis*-proline Pro241 which is found on the loop after β-strand $\beta 7$. In xylanse A the corresponding loop is much longer and has a different conformation.

The non-proline *cis*-residue, Ala80, corresponds to Thr81 in Cex which is also in the *cis*-conformation. A



(*a*)



(*b*)

Fig. 7 (*a*) Stereoview of hydrogen-bonding network of water molecules in the active site of xylanase A. A network of water molecules links the nucleophile Glu246, the acid–base Glu127 and active-site residues Gln213, His79, Asn126 and His84. (*b*) Stereoview of final *SIGMAA*-weighted (Read, 1986) $2F_o - F_c$ map (1σ contour level) at 1.8 Å resolution showing the well ordered water molecules at the active site of xylanase A. The nucleophile Glu246 and the acid–base Glu127 are shielded by aromatic residues. [Displayed using the graphics program *FRODO* (Jones, 1978).]

map from a simulated-annealing *X-PLOR* run in which the region around Ala80 has been omitted (Hodel, Kim & Brünger, 1992) is shown in Fig. 5. This residue is usually, though not invariably, a threonine in family F xylanase sequences. However, the preceding histidine residue (His79 in xylanase A, His80 in Cex) is highly conserved. This is at the end of β-strand β3 and both structures are very similar in this region. His79 may be important as it is close to the active site, and together with His77, helps orient Asp123 on β-strand β4 (Fig. 6*c*). All these three residues are highly conserved across the family and mutational studies implicate Asp123 in a direct or indirect role in catalysis (MacLeod *et al.*, 1994). White *et al.* (1994) rationalize the *cis*-conformation of Thr81 in Cex in terms of rotation necessitated to avoid a steric clash with the carbonyl of Asp123. In xylanase A *cis*-Ala80 is stablized by hydrogen bonds with two water molecules (OW353 and OW367). The solvent structure is also very similar in this region in Cex. Although unusual, non-proline *cis*-peptides have been observed in a number of other high-resolution refined protein structures and are almost exclusively found in regions directly involved in ligand binding, the positioning of catalytic residues or both (Herzberg & Moult, 1991). His79 and *cis*-Ala80 form part of a functionally important β-bulge feature discussed below.

3.2. *Special structural features of xylanase A*

There are three β-bulges (Richardson, Getzoff & Richardson, 1978; Chan, Hutchinson, Harris & Thornton, 1993) found on β-strands β4 (Val124 and Val125, followed by Asn126 and Glu127, Fig. 6*b*) and β7 (Thr245 and Glu246, Fig. 6*a*), the β-strands which bear the catalytic site residues Glu127 and Glu246, respectively. These residues are all highly conserved and the same β-bulge structures are observed in the Cex structure. These β-bulges serve to orient the catalytically important residues (Glu246, Glu127, Asn126) in the active site. A survey of β-bulges in other proteins has shown that β-bulges play an important biological role by positioning crucial residues in the protein and are well conserved (Chan *et al.*, 1993).

The double β-bulge on β-strand β4 gives rise to a crankshaft-like structural feature involving the residues Val124, Val125, Asn126, Glu127 in $\alpha_R - \beta - \beta - \alpha_L$ conformations (Fig. 6*b*). This has the effect of altering the direction of β-strand, and then changing it back again. The crankshaft feature arises from the need to place inside two adjacent hydrophobic valine residues (Val124 and Val125) in the middle of β-strand 4; the following two polar residues, Asn126 and Glu127, are thrown outwards, giving rise to the double β-bulge or crankshaft feature.

The β-strand β3, which is adjacent to β-strand β4, itself has a β-bulge (His79 and *cis*-Ala80) followed by a β-turn at the end of the distorted β-strand involving residues Ala80, Leu81, Val82 and Trp83 (Fig. 6*c*), the distance between residue i and $i+3$ of 6.7 Å (< 7 Å) and the conformation of residue $i+1$ and $i+2$ make this a β-turn of the type $\alpha_R - \beta_E$ or VIII (Wilmot & Thornton, 1990). This is called a β-turn rather than a β-bend because it is at the end of a β-strand. The double β-bulge β-turn feature has the effect of orienting His79 in the active site and in this way serves the same function as the β-bulges on β-strands β6 and β4. The β-bulge (Val124 and Val125) on strand β4 is spatially adjacent to the β-turn or distorted β-bulge involving Leu81 and Val82. The β-bulge β-turn feature is also present in the Cex structure, where nearly all the residues involved with the exception of *cis*-Ala80 which is *cis*-Thr81 in Cex, are conserved. The structural conservation of these β-bulges is consistent with them being of functional importance.

His79 is preceded by a glycine (Gly78) allowing greater flexibility of the β-strand before the β-bulge β-turn distortion. This glycine is highly conserved across family F xylanases. The space under Gly78 is filled by the side chain of Trp122 on the adjacent β-strand β4 (Fig. 6*c*). This tryptophan residue is correspondingly present in Cex and is an aromatic residue (usually a tryptophan, sometimes a phenylalanine) in family F xylanases.

The active-site residues Asn126 and the acid–base catalyst Glu127 on β-strand 4, and the corresponding pair of residues on the adjacent β-strand 5, Asn179 and Asp180, are highly conserved across family F xylanases. Asn126 and Glu127 are on a β-bulge, so both point into the active site, as does Asn179. The position of Asn179 enables hydrogen bonding to both Asn126 and the nucleophile Glu246. Asparagine is a commonly occurring residue on β-bulges (Chan *et al.*, 1993) and the occurrence of Asn126 next to the acid–base Glu127 on a β-bulge may be an important factor in the positioning of Glu127. Asn126 may also shield Glu127 and play a role in protonation and orientation of Glu127.

An extensive hydrogen-bonding network involving the catalytic residues at the active site of xylanase A is likely to play a vital role in the detailed mechanism of catalysis. The nucleophile, Glu246, is held in place by a hydrogen bond to His215, which in turn forms hydrogen bonds to Asp248. Glu246 also forms a hydrogen bond to Asn179, which forms a hydrogen bond to Asn126. Asn126 is oriented in the active site by Trp83 which forms a hydrogen bond to the acid base Glu127. The acid–base Gln127 forms a hydrogen bond to Gln213. A very similar hydrogen-bonding network involving residues of the active site is observed in Cex, as would have been expected, since all these residues are highly conserved across the family F xylanases and likely to be of importance for catalysis.

In addition to the above hydrogen-bonding network of active-site residues, in the catalytic site of xylanase A there is also an extensive hydrogen-bonding network of

water molecules linking active-site residues. In particular, a hydrogen-bonded chain of four water molecules links the acid–base Glu127 to the nucleophile Glu246 (Fig. 7). Further water molecules extend this chain to link in His84. Additional water molecules link the chain to other residues in the active site such as His79 and Gln213, reinforcing the hydrogen-bonding pattern linking the active-site residues (Fig. 7*a*). These water molecules are not present in the Cex structure, which has only 45 water molecules. The criteria for selection of water molecules in Cex (White *et al.*, 1994), namely, within hydrogen-bonding distance of 2.5–3.5 Å to at least two protein atoms, would probably result in the exclusion of most of these water molecules which form either only one or no hydrogen bonds to any protein atoms.

In xylanase A, a chain of 20 well ordered water molecules runs along the length of the substrate-binding cleft. Substrate-binding studies of the inactive xylanase A E246C mutant (Harris *et al.*, 1994) have shown that the xylopentaose substrate is bound in the cleft in such a way that some or all of these water molecules are expelled on binding of the substrate.

In xylanase A the active-site residues are boxed in by aromatic residues, Trp83, Trp305, Trp313 and Phe181 (Fig. 7*b*). All these aromatic residues are conserved across family F xylanases, with the exception of Phe181, which is a tyrosine residue in the corresponding position in other family F xylanases. In carbohydrate-binding proteins, specificity and stability is conferred by partial stacking of aromatic residues with the sugar ring structure (Quiocho, 1986). In xylanase A, the conserved tryptophan residues Trp83, Trp305 and Trp313, Phe181 and two tyrosine residues that are not correspondingly present in other family F xylanases, Tyr220 and Tyr255 (the latter is on the long loop after β-strand β7) line the substrate-binding cleft and make contacts with the xylopentaose substrate (Harris *et al.*, 1994). These residues are important in substrate binding and could also play a role in substrate specificity.

3.3. *Crystal packing*

There are four copies of the xylanase A molecule in the asymmetric unit and the packing in the crystals is pseudo-tetragonal. These four molecules are essentially the same. The r.m.s. deviations between molecules 1 and 2, 1 and 3, and 1 and 4 are 0.172, 0.225 and 0.216 Å, respectively, for 345 Cα atoms using *MNYFIT* (Sutcliffe *et al.*, 1987). There are only small differences in side-chain positions. The largest differences occur on the surface of the protein and are probably due to crystal-packing effects.

The four xylanase A molecules form two pairs of molecules (molecules 1 and 2, and molecules 3 and 4). Within a pair, the two molecules are related by an approximate twofold axis. The relationship between molecules 1 and 2 is the same as that between molecules 3 and 4, but the twofold axes are not along the same direction. Molecules 1 and 2 are related by an approximate twofold axis in or close to the *yz* plane, at 45° to the *y* axis. Molecules 3 and 4 are related by an approximate twofold axis close to the *xy* plane, at approximately 45° from the *x* axis.

The interface between molecules 1 and 2 involves contacts between the loops at the C-terminal ends of the β-strands. Part of the long loop after β-strand 7 (residues 264–267) in molecule 1 makes contact with the end of the loop after β-strand 1 (residues 17–19) in molecule 2. An adjacent region in the same long loop (residues 257–260) makes contacts with the end of the loop after β-strand 2 (residues 50–52) and the end of the loop after β-strand 3 (residues 86–94) in molecule 2. The latter loop in molecule 2 also makes contacts with the end of the loop after β-strand 8 (residues 318–320) in molecule 1. The loop after β-strand 4 in molecule 1 (residues 133–138) interacts with the corresponding loop in molecule 2 (residues 138 to 133) in a pseudo-symmetric manner about the approximate twofold axis relating molecule 1 to molecule 2. The twofold axis also serves to repeat the intermolecular interactions listed above for molecules 1 and 2 and these same interactions are found between the corresponding loops in molecules 2 and 1. The molecular interface contacts between molecules 3 and 4 are similar to those between molecules 1 and 2.

The structure of the xylanase A active-site mutant E246C has previously been determined at 2.5 Å (Harris *et al.*, 1994). The E246C xylanase A mutant crystallizes in a tetragonal space group, $P4_32_12$, with two molecules in the asymmetric unit. The relationship between the two molecules is the same as that between the pairs of molecules (1 and 2 or 3 and 4) in the native xylanase A structure in the orthorhombic space group $P2_12_12_1$. In the tetragonal xylanase A E246C mutant structure, the two molecules are related by an approximate twofold axis in the *yz* plane, the same as that which relates molecules 1 and 2 in the orthorhombic native xylanase A structure. The pairs of molecules superimpose well. The molecular interface contacts are also similar. However, there is no evidence to suggest that xylanase A is a dimer in solution.

4. Concluding remarks

Xylanase A is an eightfold β/α-barrel with the major axis running from β-strand 1 to β-strand 5. The active-site glutamates, Glu127 and Glu246, are both situated on β-bulges on β-strands 4 and 7, respectively. An additional β-bulge distortion on β-strand 3 incorporates a non-proline *cis*-peptide residue (Ala80). These three β-bulge type distortions are functionally significant as they serve to orient important active-site residues. The active-site residues are held in place by an extensive hydrogen-bonding network of active-site residues. A chain of well

ordered water molecules occupies the substrate-binding cleft, some or all of which are expelled on binding of the substrate.

Coordinates of xylanase A have been deposited with the Brookhaven Protein Data Bank.*

We thank Professor N. Sakabe for assistance and generous provison of synchrotron time on BL6A at the Photon Factory, KEK National Laboratory for High Energy Physics, Japan. We also thank Geoffrey Hazlewood and Harry Gilbert for supplying the initial *xynA* gene and Mandy Scott for growing the xylanase A crystals. We are grateful to Rutherford Appleton Laboratory for time on the Cray YMP8. This work was funded by the BBSRC.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1CLX, R1CLXSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: HE0136).

References

Arase, A., Yomo, T., Urabe, I., Hato, Y., Katsube, Y. & Okada, H. (1993). *FEBS Lett.* **2**, 123–127.

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

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tsumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.

Campell, R. L., Rose, D. R., Wakarchuk, W. W., To, R., Sung, W. & Yaguchi, M. (1993). *Proceedings of the second TRICEL symposium on Trichoderma reesei cellulases and other hydrolases*, Vol. 8, edited by P. Suominen & T. Reinikainen, pp. 63–72. Helsinki: Foundation for Biotechnical and Industrial Fermentation Research.

Chan, A. W. E., Hutchinson, E. G., Harris, D. & Thornton, J. M. (1993). *Protein Sci.* **2**, 1574–1590.

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

Dauter, Z. (1993). *X-ray structures of the seed protein narbonin and of xylanase from Bacillus pumilus*, in *Proceedings of the 3rd AFRC Protein Engineering Conference*, Cambridge, England.

Derewenda, U., Swenson, L., Green, R., Wei, Y., 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., Chauvaux, S., Beguin, P. & Alzari, P. M. (1995). *Nature Struct. Biol.* **2**, 569–576.

Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.

Farber, G. K. (1993). *Curr. Opin. Struct. Biol.* **3**, 409–412.

Farber, G. K. & Petsko, G. A. (1990). *Trends Biochem. Sci.* **15**, 228–234.

Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. Jr & Warren, A J. (1991). *Microbial Rev.* **55**, 303–315.

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

Henrissat, B. (1991). *Biochem. J.* **280**, 309–316.

Herzberg, O. & Moult, J. (1991). *Proteins*, **11**, 223–229.

Hodel, A., Kim, S.-H. & Brünger, A. T. (1992). *Acta Cryst.* A**48**, 851–858.

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

Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

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

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

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

Lee, Y.-E., Lowe, S. E., Henrissat, B. & Zeikus, J. G. (1993). *J. Bacteriol.* **175**, 5890–5898.

MacLeod, A. M., Lindhorst, T., Withers, S. G. & Warren, R. A. (1994). *Biochemistry*, **33**, 6371–6376.

Molecular Simulations Inc. (1994). *Quanta*. Molecular Simulations Inc., 200 Fifth Avenue, Waltham, MA02154, USA.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Otwinowski, Z. (1993). *Oscillation data reduction program*, in *Proceedings of the CCP4 Study Weekend*, pp. 56–62, 29–30 January 1993. Warrington: SERC Daresbury Laboratory.

Pickersgill, R. W., Jenkins, J. A., Scott, M., Connerton, I., Hazlewood, G. P. & Gilbert, H. J. (1993). *J. Mol. Biol.* **229**, 246–248.

Quiocho, F. A. (1986). *Annu. Rev. Biochem.* **55**, 287–315.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Richardson, J. S., Getzoff, E. D. & Richardson, D. C. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 2574–2578.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C. & Jones, T. A. (1990). *Science*, **249**, 380–386.

Sakabe, N. (1991). *Nucl. Instrum. Methods Phys. Res. A*, **303**, 448–463.

Sutcliffe, M. J., Haneef, I., Carney, D. & Blundell, T. L. (1987). *Protein Eng.* **1**, 377–384.

Törronen, A., Harkki, A. & Rouvinen, J. (1994). *EMBO J.* **13**, 2493–2501.

Törronen, A. & Rouvinen, J. (1995). *Biochemistry*, **32**, 847–856.

Tull, D. & Withers, S. G. (1994). *Biochemistry*, **33**, 6363–6370.

Wakarchuk, W. W., Cambell, R. L., Sung, W. L., Davoodi, J. & Yaguchi, M. (1994). *Protein Sci.* **3**, 467–475.

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

Wilmot, C. M. & Thornton, J. M. (1990). *Protein Eng.* **3**, 479–493.