Protein Science (1998), 7:2081–2088. Cambridge University Press. Printed in the USA.

Structure of the xylanase from *Penicillium simplicissimum*

ANDREA SCHMIDT,[1,4] ANTON SCHLACHER,[2,4] WALTER STEINER,[3] HELMUT SCHWAB,[2] AND CHRISTOPH KRATKY[1]

[1]Abteilung für Strukturbiologie, Institut für Physikalische Chemie, Karl-Franzens Universität Graz, Heinrichstraße 28, A-8010 Graz, Austria
[2]Arbeitsgruppe Genetik, Institut für Biotechnologie, Technische Universität Graz, Petersgasse 12, A-8010 Graz, Austria
[3]Arbeitsgruppe Enzymtechnologie, Institut für Biotechnologie, Technische Universität Graz, Petersgasse 12, A-8010 Graz, Austria
[4]Spezialforschungsbereich Biokatalyse, Technische Universität Graz, Stremayrgasse 16, A-8010 Graz, Austria

(RECEIVED April 17, 1998; ACCEPTED June 2, 1998)

Abstract

Despite its relatively low pH and temperature optimum, the xylanase from *Penicillium simplicissimum* performs exceedingly well under conditions of paper bleaching. We have purified and characterized this enzyme, which belongs to family 10 of glycosyl hydrolases. Its gene was cloned, and the sequence of the protein was deduced from the nucleotide sequence. The xylanase was crystallized from ammonium sulfate at pH 8.4, and X-ray data were collected at cryo-temperature to a crystallographic resolution of 1.75 Å. The crystal structure was solved by molecular replacement using the catalytic domain of the *Clostridium thermocellum* xylanase as a search model, and refined to a residual of $R = 20\%$ ($R_{free} = 23\%$) for data between 10 and 1.75 Å. The xylanase folds in an $(\alpha/\beta)_8$ barrel (TIM-barrel), with additional helices and loops arranged at the "top" forming the active site cleft. In its overall shape, the *P. simplicissimum* xylanase structure is similar to other family 10 xylanases, but its active site cleft is much shallower and wider. This probably accounts for the differences in catalysis and in the mode of action of this enzyme. Three glycerol molecules were observed to bind within the active site groove, one of which interacts directly with the catalytic glutamate residues. It appears that they occupy putative xylose binding subsites.

Keywords: crystallization; family 10 xylanase; purification; sequencing; TIM barrel; X-ray structure

Xylanases (E.C. 3.2.1.8) degrade xylan in plant cell walls while leaving the cellulose fabric widely intact. This class of enzymes has found several biotechnological applications, specifically in pulp and paper processing, where they can be used as bleaching agents or boosters, helping to reduce the amount of chemicals required to achieve an appropriate white level. The frequently observed stability of xylanases with respect to pH and temperature under the conditions of pulp and paper bleaching invites such industrial applications. Recently, several xylanases from different families and with different molecular characteristics were compared with respect to their performance in the bleaching process (Gübitz et al., 1997). No correlation between the increase in brightness as the most relevant parameter for the bleaching process and any one of several molecular or biochemical characteristics (pI, molecular weight, pH, and temperature optima) became apparent.

Despite the fact that the xylanase from *Penicillium simplicissimum* shows a rather low pH optimum (5.6) and a moderate temperature optimum (67 °C), it performed exceedingly well in the bleaching process. This might be due to a specific mode of action of this enzyme, which differed from other tested hydrolases: during xylanase treatment, the fluidity of the reaction mixture was continuously monitored, and it showed the most rapid increase for the reaction medium with the *P. simplicissimum* enzyme. This could be due to a specific mode of xylan degradation by this enzyme, which would preferably cleave the polysaccharide chain in the middle and not near its ends (Gübitz et al., 1997). Although the biochemical properties of the studied xylanases could not explain this peculiar behavior during catalysis, we determined the three-dimenisonal structure of this enzyme by X-ray crystallography to find molecular reasons for its apparent efficiency in xylan degradation. Prior to the structure analysis, the purification protocol for this protein was optimized and the primary structure was determined by cloning and sequencing its gene.

Two types of xylanases exist on the basis of sequence similarities (Henrissat, 1991; Henrissat & Bairoch, 1993), differing in molecular weight as well as in the general architecture of the molecule (Davies & Henrissat, 1995). In the present communication, we describe the results of a crystal structure analysis of the xylanase from *P. simplicissimum*, structurally belonging to family 10 (formerly family F) of glycosyl hydrolases (Henrissat & Bairoch, 1993). To our knowledge, this is the first structure analysis of a "stand-alone" family 10 xylanase, i.e., a xylanase that is not connected to an extra cellulose binding domain. All previous reports of crystallographic structure determinations of family 10 xy-

Reprint requests to: C. Kratky, Institut für Physikalische Chemie, Karl-Franzens Universität, Heinrichstraße 28, A-8010 Graz, Austria; e-mail: christoph.kratky@kfunigraz.ac.at.

lanases concerned enzymes whose cellulose binding was either deleted from the gene or enzymatically cleaved off the mature protein. This was the case with the enzymes from *Streptomyces lividans* (Derewenda et al., 1994), *Pseudomonas fluorescens* (Harris et al., 1994, 1996), *Cellulomonas fimi* (White et al., 1994), and *Clostridium thermocellum* (Dominguez et al., 1995).

Results and discussion

Purification, cloning, and sequencing

We have purified the xylanase from *P. simplicissimum* following a three-step protocol consisting of ammonium sulfate precipitation, hydrophobic interaction chromatography, and ion exchange chromatography.

The genomic sequence of the xylanase was determined as follows: primers were designed according to the sequences of highly conserved regions within xylanases of related fungi. With these primers, specific PCR products could be obtained that were used to screen a library of *Hind*III 4.5 and 6.5 kb fragments. A proper clone could be isolated and yielded the sequence of the C-terminal part of the xylanase. The missing N-terminal part was then obtained by specifically designed primers based on homologies in the signal sequences of related fungal xylanases. The entire mature protein is encoded within 1,370 basepairs of genomic sequence; 9 exons encoding 302 amino acids of the mature xylanase protein are interrupted by 8 introns, as identified by sequence alignment and analysis of splice site consensus sequences. The deduced amino acid sequence, as expected, revealed high homology to the sequences of the *Penicillium chrysogenum* and *Aspergillus awamori*, as well as other family 10 xylanases (Fig. 1)

Crystallization and X-ray analysis

The protein was crystallized at 4 °C as described in Materials and methods. X-ray diffraction data were collected at cryo-temperature (resolution range of 10 to 1.75 Å), the structure was solved by molecular replacement using the *C. thermocellum* xylanase catalytic core (Dominguez et al., 1995) as a search model, and refined to a crystallographic residual of $R = 20\%$ ($R_{free} = 23\%$).

The electron density was found to be well defined throughout most of the protein chain (Fig. 2), except for the Lys232 side chain, which appears to be dynamically disordered and produces no visible density. Trp276 forms a "lid" over the active site and is presumably mobile, resulting in weaker density. Three side chains were refined in two alternate conformations: Ser90, Arg155, and Ser163 (50%/50% for Ser90 and Ser163, 65%/35% for Arg155). Ser90 is located near the catalytic center, and probably participates in substrate binding. There is one disulfide bridge connecting residues 256 and 262 and fixing the intervening residues into a turn conformation, similar to the structure of the *C. fimi* xylanase domain.

The N-terminal Gln side chain is modified into a pyro-form by ring formation with its α-NH_2 group, resulting in a lactame ring that was clearly visible in the electron density. N and C termini are in close proximity, and their electron density is well defined. No disorder was observed in the loop regions on the "top" of the barrel. Ninety-one percent of the non-Gly and non-Pro residues appear in the core regions of a Ramachandran plot.

Molecular architecture

The xylanase from *P. simplicissimum* folds into an $(\alpha/\beta)_8$ or TIM-barrel (Banner et al., 1975) with an elliptical cross section (Fig. 3). The "top" of the barrel is widened, resulting in a 50 × 30 Å wide

```
Psimp   ----------------------------QASVSIDAKFKAHGKKYLGTIGDQYTLt--------KNtKnPAIIKADFGQ   43
Pench   mipnitqlktaalvmlfagqalsgpvesrQASESIDAKFKAHGKKYLGNIADQGTLn--------gNpKTPAIIKAnFGQ   72
Aspak   mvqikaaalamlfashvlsepiepr----QASVSIDSKFKAHGKKYLGNIGDQYTLt--------KNsKTPAvIKADFGa   68
Clotm   sgnalrdyaeargikigtcvnypfynn-------------------------------------sdptynsIlqreFsm   42
Celfi   mprttpapghpargartalrttrrraatlvvgatvvlpaqaattlkeaadgagrdfgfaldpnrlseaqykAIadseFnl   80

Psimp   LTPENSMKWDATEPnRGQFtFSGSDYLVNFAQSNGKLIRGHTLVWHSQLPGWVS--SITDKNTLISVLKNHITTVMTrYK  121
Pench   LsPENSMKWDATEPSqGQFSFaGSDYfVeFAetNGKLIRGHTLVWHSQLPSWVS--SITDKTTLtdVMKNHITTVMkqYK  150
Aspak   LTPENSMKWDATEPSRGQFSFSGSDYLVNFAQSNnKLIRGHTLVWHSQLPSWVq--aITDKNTLIeVMKNHITTVMqHYK  146
Clotm   vvcENEMKFDAlqPrqnvFdFSKgDqLlaFAerNGmqmRGHTLIWHNQNPSWltngnw-nrdsLlaVMKNHITTVMTHYK  121
Celfi   vvaENAMKWDATEPSqnsFSFgagDrvasyAadtGKelyGHTLVWHSQLPDWak--nl-ngsafeSaMVNHvTkVadHfe  157

Psimp   GKIYAWDVLNEIFNEDGSLRnS-VFYNVIGEDYVRIAFETARsvDPNAKLYINDYNLDSAgYsKvNGMVSHVKKWLAAGi  200
Pench   GKLYAWDVVNEIFeEDGtLRDS-VFsrVlGEDFVRIAFETAReADPeAKLYINDYNLDSAtsaKLqGMVSHVKKWIAAGV  229
Aspak   GKIYAWDVVNEIFNEDGSLRDS-VFYkVIGdDYVRIAFETARAADPNAKLYINDYNLDSAsYpKLaGMVSHVKKWIeAGi  225
Clotm   GKIveWDVaNEcmddsGnglrssiwrNVIGqDYldyAFRyAReADPdAlLfyNDYNiedl-gpKsNavfnmiKsmkerGV  200
Celfi   GKvasWDVVNEafadgdgppqdsafqqklGngYietAFRaARAADPtAKLcINDYNvegi-naKsNslydlVKdfkArGV  236

Psimp   PIDGIGSQTHL--GAGAGsAvaGALNALASAGTK---EIAITELDIA-----------GASSTDYVnVVNACLnQaKCVG  264
Pench   PIDGIGSQTHL--GAGAGAAaSGALNALASAGTe---EvAvTELDIA-----------GAtSTDYVdVVNACLDQPKCVG  293
Aspak   PIDGIGSQTHL--sAGgGAGiSGALNALAgAGTK---EIAvTELDIA-----------GASSTDYVEVVEACLDQPKCiG  289
Clotm   PIDGvGFQcHf--ingmspeylasidqnikryaeigvivsfTEiDIripqsenpatafqvqanNYkElmkiCLanPnCnt  278
Celfi   PlDcvGFQsHL--ivgqvpgdfrqnlqrfadlgv---dvrITELDIrmrtpsdatkl-atqaaDYkkVVqACmqvtrCgG  310

Psimp   ITVWGVAD-----PDSWRSSSSPLLFDgNYNPKaAYNAIANAL----------------------                302
Pench   ITVWGVAD-----PDSWRadeSPLLFDasYNPKeAYNvsqllsrqhafdlylklgnlllsrlhsd                  353
Aspak   ITVWGVAD-----PDSWRSSStPLLFDsNYNPKPAYtAIANAL----------------------                327
Clotm   fvmWGftDkytwiPgtfpgygnPLiyDsNYNPKPAYNAIkeALmgy--------------------                324
Celfi   vTVWGitDkyswvPDvfpgegaaLvwDasYakKPAYaAVmeAfgas--------------------                356
```

Fig. 1. Sequence alignment of family 10 xylanases, computed with program MACAW Ver. 2.0.0 (Schuler, 1993). Xylanases from the following organisms were included: *P. simplicissimum* (Psimp), *A. awamori* (Aspak, swissprot entry code P33559), *P. chrysogenum* (Pench, P29417), *C. thermocellum* (Clotm, P10478), and *C. fimi* (Celfi, P07986). Conserved blocks marked in yellow.

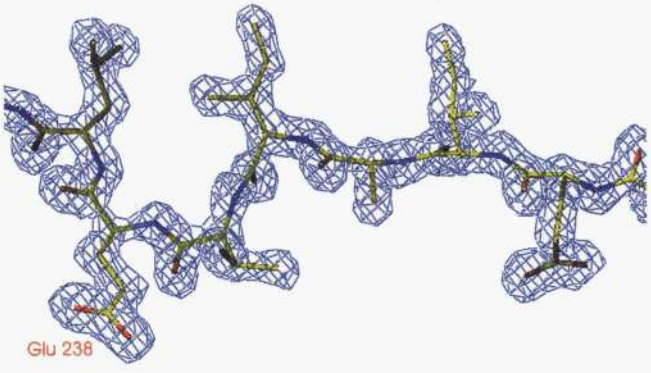

Fig. 2. Electron density ($2F_o - F_c$ map, contoured at 2σ level) originating from strand VII of the crystal structure of the *P. simplicissimum* xylanase. Glu238 is the residue proposed to act as a nucleophile during catalysis (figure produced with program O, Jones et al., 1991).

and 25 Å high "bowl." Its "rim" is higher at the end points of the long axis of the cross-section ellipse than at the end points of the short axis, which is caused by protruding helices and loops. It, therefore, seems that the upper region is more flexible compared to the bottom, where shorter connections occur between helices and strands. A similar type of folding has been observed in other family 10 xylanases, such as the catalytic domains of the *C. fimi* (White et al., 1994), *C. thermocellum* (Dominguez et al., 1995), *P. fluorescens* (Harris et al., 1994, 1996), and *S. lividans* (Derewenda et al., 1994) xylanases. Differences in the molecular structure occur in the N- and C-terminal regions and in the length of the loops and helices on the barrel top, and they may account for differences in substrate specificities and/or additional cellulase activity of these xylanases. Compared to the other structures, the *P. simplicissimum* xylanase has shorter loops and a correspondingly more compact appearance (Fig. 3).

The active site is located in a cleft running across the barrel top. The cleft's central part is very narrow and contains the catalytic residues Glu132 and Glu238. The cleft widens and becomes more shallow toward its ends. Residues near the catalytic center, for example, the two glutamate residues, Trp88, Trp268, Trp276, His84, and His210 are strictly conserved in family 10 xylanases (Fig. 1).

Prior to their crystallographic investigation, crystals of the xylanase were briefly soaked in a cryo-protectant containing 33% glycerol. As a result, electron density corresponding to seven glycerol molecules in direct contact with the protein surface were observed, three of them binding directly within the active site cleft. One glycerol is involved in close H-bonding contacts to the catalytic residues; the other two are arranged in a row along the cleft (Fig. 4). The arrangement of these three molecules therefore suggests possible binding sites for the xylose subunits of the natural substrate xylan.

Figure 4 also shows a corresponding surface representation for the xylanase from *C. fimi*. For this enyme, the crystal structure of a complex with 2-fluoro-2-deoxy-cellobioside was also determined (White et al., 1996), and Figure 4 shows the inactivator, which is covalently bound to the catalytic nucleophile (the residue equivalent to Glu238, see Fig. 1). Two of the glycerol molecules in the *P. simplicissimum* structure are observed at very similar locations as the two glucose rings in the *C. fimi* complex, with the oxygen atoms of the two glycerol molecules engaged in similar H-bonds as the glucose oxygens (Fig. 5).

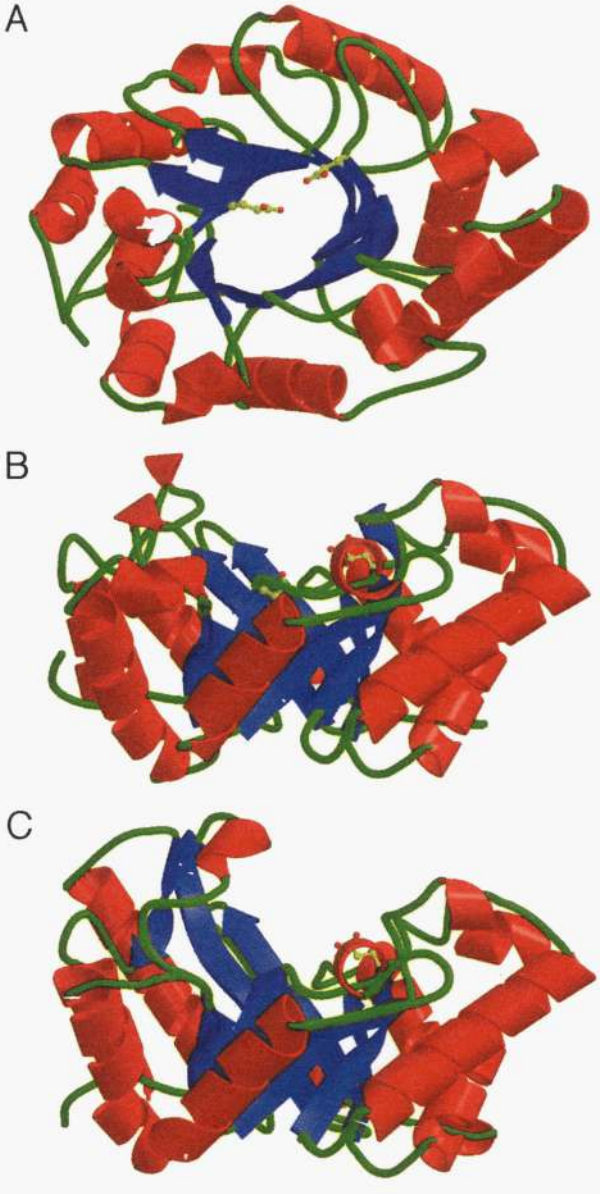

Fig. 3. Two views of the structure of the *P. simplicissimum* structure with the catalytic glutamate residues. (**A**) Top view, (**B**) side view, and (**C**) side view of the catalytic domain of the *C. fimi* xylanase. Figure drawn with MOLSCRIPT (Kraulis, 1991).

Trp276 forms a "lid" above the catalytic center, and appears to partially shield the active site. This probably forces the enzyme-bound xylan chain into a bend. Several aromatic residues are arranged along the cleft (Trp88, Trp268, Trp276, Tyr175), which simultaneously shows a negative surface charge due to the presence of hydrophilic residues (Glu47, Asn48, Gln91, Asp53, Ser90, Asn176, Asp178).

Using chemical as well as genetic techniques, the identity and function of the catalytic residues in family 10 xylanases were assigned unambiguously for the enzymes from *C. fimi* (Tull et al., 1991; MacLeod et al., 1994) and *S. lividans* (Moreau et al., 1994; Roberge et al., 1997). Alignment of the *P. simplicissimum* xylanase with these enzymes (Fig. 1) strongly suggests Glu132 to act as the proton donor and Glu238 as the nucleophile in the enzymatic

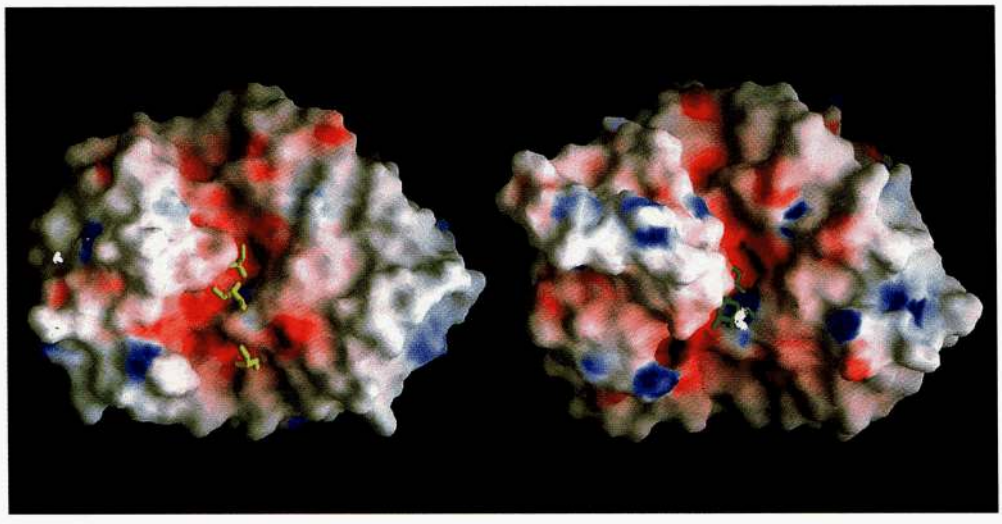

Fig. 4. Surface representation of the xylanases from *P. simplicissimum* (left) and *C. fimi* (right). Figure produced with program GRASP (Nicholls, 1993). Negatively charged surface areas are colored in red, positively charged ones in blue. Also shown are three glycerol molecules found within the active site cleft of the *P. simplicissimum* structure as well as a cellobioside inactivator covalently attached to the nucleophile of the *C. fimi* enzyme. Data for the *C. fimi* structure were taken from White et al. (1996).

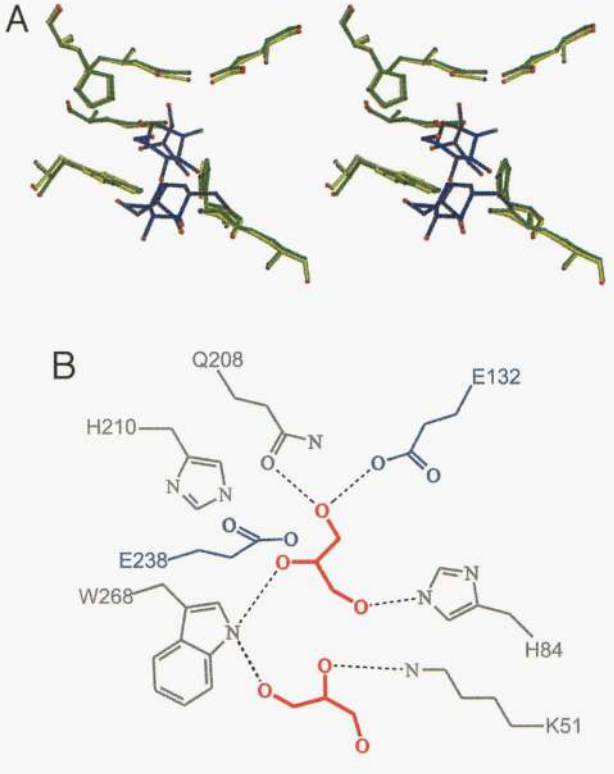

Fig. 5. A: Stereographic superposition of topologically equivalent residues in the active sites of the xylanase crystal structures of *P. simplicissimum* (yellow) and *C. fimi* (green). Also shown are two glycerol molecules observed in the *P. simplicissimum* structure and a covalently attached 2-deoxy-2-fluoro-cellobioside inhibitor present in the *C. fimi* structure. **B:** Hydrogen bonding interactions between two glycerol molecules and protein side chains as observed near the active site of the *P. simplicissimum* structure.

reaction. A conserved block of residues containing the motif Asn-Glu, with Glu acting as is the proton donor, further confirms this assignment, which is also consistent with the higher surface exposure of Glu132 compared to Glu238. The distance between the two side-chain functional groups is around ~5.5 Å, in line with retention of the xylose anomeric configuration via a double displacement mechanism with a covalently bound intermediate (McCarter & Withers, 1994; Davies & Henrissat, 1995). This active site architecture is conserved in family 10 xylanases and has also been observed in the *C. fimi, C. thermocellum, P. fluorescens,* and *S. lividans* structures. Recently, the *P. simplicissimum* xylanase has also been shown to catalyze hydrolysis upon retention of the anomeric configuration (G. Gübitz, pers. obs.).

Thr85 forms a *cis*-peptide bond (between residues His84 and Thr85), permitting a sharp bend that brings His84 into a position within the cleft where it can interact with the bound substrate. This special *cis*-peptide bond, embedded in a conserved block of residues, is also present in other family 10 xylanase structures.

Implications for biocatalysis

A comparison of the overall shape of the *P. simplicissimum* with the *C. fimi* (White et al., 1994) enzyme reveals several novel features (Figs. 3, 4). Although the shape of the active site cleft is altogether similar in the two enzymes, the *P. simplicissimum* xylanase has a shallower and wider xylan-binding groove. A similar conclusion emerges from a comparison with other known family 10 xylanase structures (Derewenda et al., 1994; Harris et al., 1994, 1996; Dominguez et al., 1995). The shallowness of its binding cleft in the *P. simplicissimum* enzyme is due to the lack of long loops on the barrel top (Fig. 3). Subsites for the binding of xylose subunits can be inferred from the charge distribution around the active site. It shows two to three charged patches for *P. simplicissimum*, but a much longer putative binding site in the *C. fimi* xylanase (Fig. 4).

In the *C. fimi* enzyme, the cleft differs in shape as well as in charge distribution. Although the *C. fimi* xylanase is known to also display exo-cellulase activity (Tull & Withers, 1994), the *P. simplicissimum* xylanase has only neglegible cellulase activity (G. Gübitz, pers. obs.).

The observation of a shallow active site cleft in the *P. simplicissimum* xylanase is somewhat surprising in view of the fact that, among the above family 10 xylanases, the *P. simplicissimum* enzyme is the only one without a covalently attached cellulose binding domain (CBD) in its native form. One might have expected that lack of a CBD necessitates higher specificity of the catalytic site, which might then result in a longer and narrower active site cleft. Although three-dimensional structural data are known from NMR spectroscopy for isolated cellulose binding domains (Xu et al., 1995; Johnson et al., 1996), we know of no such data for family 10 xylanases with intact CBDs.

It is tempting to speculate that the observed shape of its active site cleft is one of the reasons for the good performance in the pulp bleaching procedure of the *P. simplicissimum* xylanase: release of lignin from pulp is most effective if the polymer chain is first cleaved in the middle and not near one of its ends, yielding fragments of medium size. This would cause the reaction mixture to quickly decrease in viscosity (Gübitz et al., 1997), which favors diffusion and release of the lignin. Due to the unobstructed active site of the *P. simplicissimum* xylanase, binding of the xylan chain is probably less hindered than in related enzymes, and may even admit branched polysaccharides. However, appraisal of such structural inferences will have to await detailed kinetic data from a variety of glycosidic model substrates, which are so far only available for the *C. fimi* enzyme (Tull & Withers, 1994).

Materials and methods

Organism and growth conditions

Penicillium simplicissimum (Oudem.) Thom (BT 2246) was from the culture collection of the Institute of Biotechnology, Graz, University of Technology, Austria. It was originally isolated from decomposed woody material in Styria, Austria, and was identified by Centraalbureau voor Schimmelcultures (CBS), Baarn, The Netherlands. Stock cultures were maintained on Sabouraud-agar (SAB) at 4 °C and transferred every 6–7 weeks. SAB plates were incubated at 30 °C for 2–4 days. For shake flask experiments each 300-mL Erlenmeyer flask containing 100 mL medium was inoculated with a piece (1 cm^2) from a fungal colony actively growing on SAB plates and incubated at 30 °C for 7–8 days under agitation (150 rpm, 25 mm stroke) on an orbital shaker. The fungus was grown in a medium containing 4% coarse milled Corncobs, 3% Solulys (Roquette, Lille, France), 0.5% KH_2PO_4, and 0.8% NH_4NO_3 in tap water at a starting pH of 6. The culture broth was then centrifuged, and the clear supernatant was stored at 4 °C for further use after addition of Thimerosal as biocide. The yield of such a crude enzyme preparation is about 3.5 U/mL (58.4 nkat/mL).

Isolation and purification

Chemicals used for purification were purchased from Sigma (St. Louis, Missouri) and Fluka (Buchs, Switzerland), water was deionized, and further processed using a MilliQ system (Millipore, Bedford, Massachusetts). The chromatographic purification steps were carried out on a BioLogic chromatography system (Biorad, Hercules, California), for the hydrophobic interaction step a self-packed 75-mL Biorad column was prepared, using the Sephadex75 medium from Pharmacia (Uppsala, Sweden) and following the instructions of the manufacturer.

The purification of the protein was performed following a three-step protocol adapted from an existing two-step procedure (Lischnig, 1998), which involved ammonium sulfate precipitation and hydrophobic interaction chromatography (HIC). The first step of the modified protocol consists of a precipitation with $(NH_4)_2SO_4$: to 660 mL crude extract, solid $(NH_4)_2SO_4$ was added at 4 °C under constant stirring (100 rpm) to increase the concentration in steps of 10% saturation up to an end concentration of 60%. The amount of added $(NH_4)_2SO_4$ was calculated from the initial volume and the saturation at 20 °C [100% saturation correspond to 70.60 g $(NH_4)_2SO_4$ in 100 mL of solution]. To complete the precipitation, stirring was continued overnight. The end volume of the mixture was then 790 mL, giving a true precipitant concentration of 50%. The suspension was centrifuged at 3,000 rpm (Sorvall SS-34 rotor) at 4 °C for 1 h, and the pellets were resuspended in a total of 80 mL of 0.7 M $(NH_4)_2SO_4$/20 mM NaH_2PO_4 at pH 6.0 to dissolve the xylanase. The suspension was again centrifuged and yielded 84.5 mL of a dark-brown, clear solution. These steps were monitored by SDS-PAGE for the presence of the xylanase protein.

The supernatant was loaded on a phenylsepharose HIC column and eluted using a $(NH_4)_2SO_4$ gradient between 1 M (B) and 0 M (A) in 20 mM NaH_2PO_4 at pH 6.0. The elution volumes were as follows: 75 mL from 100 to 80% B, 150 mL from 80 to 20% B, 75 mL from 20 to 0% B at a flow rate of 4 mL/min at room temperature. Xylanase was eluted from the column at 0.5 M $(NH_4)_2SO_4$, the total amount of xylanase protein within the pooled fractions was ~130 mg. SDS-PAGE analysis still showed some minor bands besides the xylanase.

The total volume of the HIC eluate was first dialysed against 50 mM TrisHCl pH 7.5/25 mM $(NH_4)_2SO_4$, subsequently against 20 mM TrisHCl pH 6.8, both times at 4 °C overnight. This solution was loaded onto a Mono-S ion exchange column (Biorad, 10 mL) and eluted with a NaCl gradient from 0 M (A) to 1 M (B) at a flow rate of 3 mL/min and an elution volume of 39 mL from 0 to 10% B at room temperature. The xylanase elutes at 2% B. The final protein concentration of 2.8 mg/mL was determined both with UV absorption at 280 nm and with a BCA protein assay (Pierce, Rockford, Illinois), giving a total xylanase yield of 32 mg pure enzyme. This protein sample showed a single band on SDS-PAGE and isoelectric focusing gels (PHAST, Pharmacia) and was used for crystallization. The solution was concentrated with Millipore Ultrafree filters (cutoff 10 kDa).

Protein characterization

The molecular weight of the xylanase was determined by MALDI-TOF (Matrix Assisted Laser Desorption and Ionization–Time-of-Flight spectroscopy, performed on a Kratos Compact MALDI 2 V4.00 with sinapinic acid as matrix) to ~32,600 Da. The isoelectric point of the xylanase was determined by isoelectric focusing (PHAST, Pharmacia) to be near or slightly above pH 8.

Partial protein sequence information was obtained by a combination of MALDI-TOF and ladder-sequencing (Chait et al., 1993). Following an 18-h tryptic digest, fragments were separated by reversed-phase HPLC and characterized by MALDI-TOF. Two peptides were sequenced with an adapted ladder sequencing tech-

nique (Metzger, 1994), of which the larger one was later used to identify the correct gene during DNA sequencing.

Gene cloning and sequencing

Materials and general techniques

Chemicals used for the preparation of buffers were from Sigma and Fluka, restriction enzymes and corresponding reaction buffers from Boehringer Mannheim (Mannheim, Germany). All plasmid DNA isolation and cloning procedures as well as gel electrophoretic analyses of DNA were performed according to standard methods (Sambrook et al., 1989). For Southern blots, N+ membranes of Boehringer Mannheim or Hybond membranes of Amersham (Buckinghamshire, UK) were used. Primers for PCR and sequencing were synthesized on an Applied Biosystems (Foster City, California) 392 DNA/RNA synthesizer employing nucleotide reagents from Perkin-Elmer (Foster City, California). For sequencing, the dideoxy chain termination method of Sanger (Sanger et al., 1977) was applied, and products analyzed on an automated Applied Biosystems 373A DNA sequencer.

Isolation of fungal genomic DNA

Fungi were grown on cellophane-coated agar culture plates using Sabouraud medium. The mycelium was peeled off and ground to a fine powder under liquid nitrogen. The isolation of DNA from the tissue was performed using the Genomic Tip DNA Isolation Kit from QIAGEN (Hilden, Germany) following the instructions for maxi preparation. After the procedure, the genomic DNA was dissolved in 300 μL H_2O giving solutions of ~20 ng/μL DNA, which were used for all subsequent procedures.

PCR and primer design

PCR was used to amplify portions of the xylanase gene. For that, a set of four forward and four reverse primers was designed according to conserved regions found with other fungal xylanases, specifically *P. chrysogenum* and *A. awamori* sequences:

1f	5′ AA/GTTCAAGGCC/TCACGG/AAAG 3′	forward
2f	5′ CCG/AGAGAAC/TAGCATGAAGTGG 3′	forward
3f	5′ TGAAG/TAAC/TCACATCACCACG 3′	forward
4f	5′ AGGACGTT/CGTC/GAAT/CGAA/GATC 3′	forward
1r	3′ GTAGTTGCTG/AATG/ATTCAC/TTC 5′	reverse
2r	3′ CGT/GTTCGAT/GATGTAGTTGCT 5′	reverse
3r	3′ CAGTGGCTCGAA/GCTA/GTAA/GCG 5′	reverse
4r	3′ CAACGACTGGGCCATTTAC/TG 5′	reverse

The primers were first tested with *P. chrysogenum* genomic DNA as a template and then used for the amplification of the *P. simplicissimum* xylanase gene using all 16 primer combinations.

The PCR was carried out on a Perkin-Elmer GeneAmp PCR System 2400 thermocycler using a variety of temperature profiles. Eventually, one very well-defined product of ~1,000 base pairs length could be obtained with primer pair 2f /4r using the following cycling conditions: 1 min denaturation at 94 °C, then 30 cycles with 30 s denaturation at 94 °C, 60 s annealing at 54 °C, and 90 s extension at 72 °C. The reaction was completed by 5 min extension at 72 °C and subsequent cooling to 4 °C. The PCR product was purified with the QIAGEN Wizard PCR Prep Purification Kit, and sequenced directly using also the PCR primers. The DNA fragment was verified with the partial protein sequence information. This fragment was then used as a probe for screening southern blots and genomic libraries.

Creation of partial genomic libraries

Genomic DNA was partially digested with the following restriction enzymes and enzyme combinations: *Bam*HI, *Nco*I, *Bgl*II, *Xho*I, *Hind*III, *Pst*I, and *Hind*III/*Bam*HI, *Nco*I/*Xho*I, *Pst*I/*Xho*I, *Bgl*II/*Bam*HI. The PCR product was labeled with fluorescein according to the instruction manual of the Fluorescein Gene Images DNA labeling and detection system and used as a probe for the analysis of the genomic DNA digest on a southern blot. *Hind*III yielded two products of 4.5 and 6.5 kb, after isolation from a preparative 0.8% TAE agarose gel by agarase treatment of the gel and subsequent precipitation of the DNA. The isolated fragments were ligated into a pBluescript S/K+ vector and transformed into *Escherichia coli* SURE cells. These partial genomic libraries were screened using again the labeled PCR fragment as a probe. Positive clones were purified and plasmid DNA isolated following the BioRad Quantum Prep Plasmid Miniprep Kit protocol. Proper clones were selected after restriction analysis and sequenced using vector and xylanase specific primers. With the cloned 4.5 kb *Hind*III fragment the sequence of the C-terminal part of the xylanase could be obtained. As the N-terminus was not present in both *Hind*III partial genomic libraries, a new PCR primer set was created making use of similarities of the N-termini or the signal sequences in the other fungal xylanases, especially *P. chrysogenum*:

nterm1	5′ TCAG/AA/CTCAAGG/ACAGCTGCACTG/A 3′	forward
nterm3	5′ CGT/ACAG/AGCT/ATCT/AAGTGAGCATCGAC/T 3′	forward
reverse	5′ ATGTACAGCTTGGCGTTG 3′	reverse

The following temperature profile was applied for PCR: denaturation for 60 s at 95 °C, then 25 cycles with 45-s denaturation at 95 °C, 30-s annealing at 50 °C, and 90-s extension at 72 °C, termination by cooling to 4 °C. Products of the correct size were obtained with both primers, which completed the sequence of the xylanase gene, after purification and direct sequencing as above. The entire sequence was deposited at GenBank, accession number AF070417.

Crystallization and X-ray investigation

Crystallization

For crystallization trials, the Crystal Screen reagents of Hampton (Laguna Hills, California) were used. The crystallization was performed in hanging drops at 4 °C, using Linbro multi-well plates and siliconized glass cover slides. Reservoirs contained 750 μL of 1.9 M $(NH_4)_2SO_4$ in 0.1 M TrisHCl at pH 8.4; drops consisted of 5 μL of xylanase solution 3–6 mg/mL plus 5 μL of precipitant from the corresponding reservoir. After 3–4 days, crystals appeared that formed hexagonal rods reaching up to 0.12 mm in diameter and 0.3 to 0.8 mm in length.

All crystallographic data were collected at 100 K. For shock freezing, crystals were soaked for 30 s in a drop consisting of reservoir solution plus 33% glycerol. The crystal was then picked up with a fiber loop, dumped into liquid nitrogen, and transferred to the X-ray goniometer. The crystals belong to space group $P3_121$ with cell dimensions a = 81.02 Å and c = 113.40 Å.

X-ray investigation and data collection

A first data set was collected in-house on a Siemens rotating anode generator (CuK_α radiation, λ = 1.54 Å graphite monochromator) equipped with a MARresearch 30-cm imaging plate detector and a locally constructed cryo cooling device. Data were collected between 20 and 1.90 Å. Images were processed with the programs DENZO, SCALEPACK (Otwinowski, 1990) and TRUNCATE (Bailey, 1994). This data set (98% completeness between 20 and 2.2 Å, R_{merge} = 16.5% for 22,079 unique reflections) was used for structure solution and initial refinement.

High resolution data were subsequently collected at beamline 5.2 R at the ELETTRA Synchrotron Light Source in Trieste (λ = 0.86 Å) using a 18-cm MARresearch imaging plate detector and an Oxford Cryosystems cooling device. These data (100% complete between 10 and 1.75 Å, R_{merge} = 5.9% for 43,712 unique reflections, redundancy 4–6) were used for structure refinement.

Structure solution

Its size (32,556 Da calculated from the amino acid sequence) and a sequence comparison with other xylanases (BLAST; Altschul et al., 1990) indicated that the *P. simplicissimum* xylanase belongs to Family 10 (Henrissat, 1991). From the four structures of xylanase catalytic domains of this type in the Brookhaven Protein Database, the one from *C. thermocellum* (PDB entry code 1xyz, chain A; Dominguez et al., 1995) was chosen as a starting model (identity 45%, similarity 66%). After removing the first 31 amino acids at the N-terminus (see Fig. 1), molecular replacement was carried out with X-PLOR (Brünger, 1990).

Both a rotation search (resolution range 10–3 Å), and a subsequent translation search (10–3.5 Å) yielded one unique solution, which gave an *R*-factor of 48% (correlation F_oF_c: 33%) after rigid body refinement (8–3.5 Å). A $2F_o - F_c$ map (data to 3 Å resolution) revealed the conserved regions of the molecule, specifically the central β-strands. Ten percent of the data were set aside for cross-validation using R_{free} (Brünger, 1992).

The initial solution from molecular replacement was subsequently subjected to a sequence of graphic modeling steps using the program O (Jones et al., 1991), typically followed by positional refinement (Brünger et al., 1987) and/or a simulated annealing run (Brünger et al., 1990). These graphic remodeling steps included mutation of wrong side chains, removal or rebuilding of loop regions, modeling of missing N-terminal residues, gradual increase in resolution to 1.9 Å, inclusion of 316 water molecules, 6 sodium ions, 2 Tris, and 7 glycerol molecules, and modification of the N-terminal glutamine into pyro-glutamine. Refinement of individual *B*-factors gave a final *R*-factor of 20.0% (no σ-cutoff, corr = 94%, R_{free} = 22.9%, $corr_{free}$ = 91%) for 43,712 reflections between 10–1.75 Å and 2,289 protein nonhydrogen atoms plus 379 nonhydrogen solvent atoms (water, glycerol, Tris, and sodium). The coordinates have been deposited in the Brookhaven Protein database, entry code 1bg4.

Structure validation

The structure was validated using the program PROCHECK (Laskowski et al., 1993). A Ramachandran plot showed 91% of the non-Gly and non-Pro residues in the "core" regions, 8% in "allowed" regions, and one residue (0.4%) in the "generally allowed" regions.

Refinement with ARP/REFMAC

In parallel, the structure was refined with ARP/REFMAC (Lamzin & Wilson, 1993, 1997) starting with the first crude molecular replacement model and using all available data from 10 to 1.75 Å without σ-cutoff. The model was first rebuilt using unrestrained ARP to replace and find atoms and automatic tracing using the program PEPTIDE (Lamzin & Wilson, 1997) iteratively. After 150 cycles, the protein chain was essentially complete except for the *cis*-peptide region around His84 and Thr85. At this point, the procedure was changed to restrained refinement, using PROTIN to set up the geometrical restraints. Side chains were also built automatically by the as-yet unpublished programs DOCK/SIDE (Perrakis & Lamzin, pers. obs.). Automatic generation of side chains was successful for more than two-thirds of the residues; the remaining ones were fitted manually with program O. The model was further refined and solvent molecules added, bringing the *R*-factor down to 17.6% compared to 20.0% with X-PLOR. The two models are essentially identical; the only differences occur in regions of weak electron density and in the solvent region, where an additional glycerol molecule was observed.

Acknowledgments

We thank Thomas Lischnig for help in the purification of the enzyme and supply of the crude extract, Georg Gübitz for helpful discussions, Fritz Andreae and his coworkers for performing the MALDI-TOF and ladder sequencing analysis, and Victor Lamzin and Perrakis Anastassis for their help in setting up the automated refinement. We acknowledge the financial support from the Österreichischer Fonds zur Förderung der wissenschaftlichen Forschung through the Spezialforschungsbereich Biokatalyse as well as through project 11599. Synchrotron data were collected at the beamline 5.2 R at the Elettra Synchrotron Light Source in Trieste (Italy).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol 215*:403–410.

Bailey S. 1994. The CCP4 Suite—Programs for protein crystallography. *Acta Crystallogr D50*:760–763.

Banner DW, Bloomer AC, Petsko GA, Phillips DC, Pogson CI, Wilson IA, Corran PH, Furth AJ, Milman JD, Offord RE, Proddle JD, Waley SG. 1975. Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5A resolution. *Nature 255*:609–614.

Brünger AT. 1990. Extension of molecular replacement: A new search strategy based on Patterson correlation refinement. *Acta Crystallogr A46*:46–57.

Brünger AT. 1992. The free *R*-value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature 355*:472–474.

Brünger AT, Krukowski A, Erickson JW. 1990. Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr A46*: 585–593.

Brünger AT, Kuriyan J, Karplus M. 1987. Crystallographic *R* factor refinement by molecular dynamics. *Science 235*:458–460.

Chait BT, Wang R, Beavis RC, Kent SBH. 1993. Protein ladder sequencing. *Science 262*:89–92.

Davies G, Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure 3*:853–859.

Derewenda U, Swenson L, Green R, Wei YY, Morosoli R, Shareck F, Kluepfel D, Derewenda ZS. 1994. Crystal-structure, at 2.6-Angstrom resolution, of the *Streptomyces lividans* xylanase-A, a member of the F-family of beta-1,4-D-glycanases. *J Biol Chem 269*:20811–20814.

Dominguez R, Souchon H, Spinelli S, Dauter Z, Wilson KS, Chauvaux S, Beguin P, Alzari PM. 1995. A common protein fold and similar active-site in 2 distinct families of beta-glycanases. *Nat Struct Biol 2*:569–576.

Gübitz GM, Haltrich D, Latal B, Steiner W. 1997. Mode of depolymerization of hemicellulose by various mannanases and xylanases in relation to their ability to bleach softwood pulp. *Appl Microbiol Biotechnol 47*:658–662.

Harris GW, Jenkins JA, Connerton I, Cummings N, Loleggio L, Scott M, Hazlewood GP, Laurie JI, Gilbert HJ, Pickersgill RW. 1994. Structure of the catalytic core of the family F-xylanase from *Pseudomonas fluorescens* and identification of the xylopentaose-binding sites. *Structure* 2:1107–1116.

Harris GW, Jenkins JA, Connerton I, Pickersgill RW. 1996. Refined crystal-structure of the catalytic domain of xylanase-A from *Pseudomonas fluorescens* at 1.8 angstrom resolution. *Acta Crystallogr D52*:393–401.

Henrissat B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J 280*:309–316.

Henrissat B, Bairoch A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J 293*:781–788.

Johnson PE, Joshi MD, Tomme P, Kilburn DG, Mcintosh LP. 1996. Structure of the N terminal cellulose binding domain of *Cellulomonas fimi* cenc determined by nuclear magnetic resonance spectroscopy. *Biochemistry 35*:14381–14394.

Jones TA, Zou JY, Cowan S, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A47*:110–119.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Lamzin VS, Wilson K. 1993. Automated refinement of protein models. *Acta Crystallogr D49*:129–147.

Lamzin VS, Wilson K. 1997. Automated refinement for protein crystallography. *Methods Enzymol 277*:269–305.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK version 2.0. Programs to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Lischnig T. 1998. Xylanases from *Penicillium simplicissimum* [Doctoral Thesis]. Technical University of Graz.

MacLeod AM, Lindhorst T, Withers SG, Warren RA. 1994. The acid/base catalyst in the exoglucanase/xylanase from *Cellulomonas fimi* is glutamic acid 127: Evidence from detailed kinetic studies of mutants. *Biochemistry 33*:6371–6376.

McCarter JD, Withers SG. 1994. Mechanisms of enzymatic glycoside hydrolysis. *Curr Opin Struct Biol 4*:885–892.

Metzger JW. 1994. Ladder sequencing of peptides and proteins—A combination of Edman degradation and mass spectrometry. *Angew Chem Int Ed 33*:723–725.

Moreau A, Roberge M, Manin C, Shareck F, Kuepfel D, Morosoli R. 1994. Identification of two acidic residues involved in the catalysis of xylanase A from *Streptomyces lividans*. *Biochem J 302*:291–295.

Nicholls AJ. 1993. *GRASP: Graphical representation and analysis of surface properties*. New York: Columbia University.

Otwinowski Z. 1990. *DENZO data processing package*. New Haven, Connecticut: Yale University.

Roberge M, Dupont C, Morosoli R, Shareck F, Kluepfel D. 1997. Asparagine-127 of xylanase-A from *Streptomyces lividans*, a key residue in glycosyl hydrolases of superfamily-4/7—Kinetic evidence for its involvement in stabilization of the catalytic intermediate. *Protein Eng 10*:399–403.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning—A laboratory manual*, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74*:5463–5467.

Schuler G. 1993. MACAW multiple alignment construction & analysis workbench 2.0.0 ed. Bethesda, Maryland: National Center for Biotechnology Information (NCBI).

Tull D, Withers SG. 1994. Mechanisms of cellulases and xylanases: A detailed kinetic study of the Exo-β-1,4-glycanase from *Cellulomonas fimi*. *Biochemistry 33*:6363–6370.

Tull D, Withers SG, Gilkes NR, Kilburn DG, Warren RAJ, Aebersold R. 1991. Glutamic acid 274 is the nucleophile in the active site of a retaining exoglucanase from *Cellulomonas fimi*. *J Biol Chem 266*:5621–5625.

White A, Tull D, Johns K, Withers SG, Rose DR. 1996. Crystallographic observation of a covalent catalytic intermediate in a beta-glycosidase. *Nat Struct Biol 3*:149–154.

White A, Withers SG, Gilkes NR, Rose DR. 1994. Crystal structure of the catalytic domain of the beta-1,4-glycanase Cex from *Cellulomonas fimi*. *Biochemistry 33*:12546–12552.

Xu GY, Ong E, Gilkes NR, Kilburn DG, Muhandiram DR, Harrisbrandts M, Carver JP, Kay LE, Harvey TS. 1995. Solution structure of a cellulose binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry 34*:6993–7009.