# First Crystallographic Structure of a Xylanase from Glycoside Hydrolase Family 5: Implications for Catalysis[†,‡]

Steven B. Larson,[§] John Day,[§] Ana Paulina Barba de la Rosa,[§,||] Noel T. Keen,[⊥,@] and Alexander McPherson*[,§]

*Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697-3900, and Department of Plant Pathology, University of California, Riverside, California 92521*

*Received January 24, 2003; Revised Manuscript Received May 9, 2003*

ABSTRACT: The room-temperature structure of xylanase (EC 3.2.1.8) from the bacterial plant pathogen *Erwinia chrysanthemi* expressed in *Escherichia coli*, a 45 kDa, 413-amino acid protein belonging to glycoside hydrolase family 5, has been determined by multiple isomorphous replacement and refined to a resolution of 1.42 Å. This represents the first structure of a xylanase not belonging to either glycoside hydrolase family 10 or family 11. The enzyme is composed of two domains similar to most family 10 xylanases and the α-amylases. The catalytic domain (residues 46−315) has a $(\beta/\alpha)_8$-barrel motif with a binding cleft along the C-terminal side of the $\beta$-barrel. The catalytic residues, Glu165 and Glu253, determined by correspondence to other family 5 and family 10 glycoside hydrolases, lie inside this cleft on the C-terminal ends of $\beta$-strands 4 and 7, respectively, with an $O_\epsilon2\cdots O_\epsilon1$ distance of 4.22 Å. The smaller domain (residues 31−43 and 323−413) has a $\beta_9$-barrel motif with five of the strands interfacing with α-helices 7 and 8 of the catalytic domain. The first 13 N-terminal residues form one $\beta$-strand of this domain. Residues 44, 45, and 316−322 form the linkers between this domain and the catalytic domain.

Glycoside hydrolases are enzymes that cleave polysaccharides at the glycosidic linkage between furanose or pyranose rings by either a configuration-retaining or a configuration-inverting acid-catalyzed mechanism. They have been classified into 91 family groupings based on sequence similarity (*1*−*9*).

Xylanases are glycoside hydrolases with specific activity toward xylan, the hemicellulosic fraction of plant tissues composing one-third of all renewable organic carbon available on earth (*10*). Study of xylanases in an effort to understand the basis of catalysis and specificity, therefore, has a twofold purpose: (1) to engineer more efficient enzymes for utilization in industrial processes that take advantage of this plentiful, renewable resource and (2) to develop inhibitors to protect this resource and economically important crops from attack by microorganisms that utilize these enzymes to destroy the same.

Xylanases, like all $\beta$-1,4-glycan hydrolases, utilize a general acid mechanism of catalysis promoted by two acidic amino acid residues, an acid or base and a nucleophile. In the case of xylanases, only glutamic acid residues have been identified as the catalytic amino acids. Specific recognition of substrate depends on the active site topology and a series of binding subsites on the surface of the enzyme (*7*−*9*).

Xylanases have been placed predominantly in families 5, 10, 11, and 43 in the glycoside hydrolase (GH)[1] family classification scheme (*1*−*3*). However, only three-dimensional structures of xylanases from families 10 and 11 have been reported. Glycoside hydrolase families 5 (GH-5) and 10 (GH-10) are members of a superfamily or clan GH-A, which also includes GH families 1, 2, 17, 26, 30, 35, 39, 42, 51, 53, 59, 72, 79, and 86. The clan concept demonstrates a broader relationship between GH families that suggests a more distant common evolutionary ancestor (*11*).

The family 11 xylanases are low-molecular weight proteins of 184−207 amino acids that fold into a $\beta$-sheet motif. Structures have been reported for enzymes from bacteria [e.g., *Bacillus circulans* (*12*−*15*), *B. circulans* strain D3 (*16*), and *Bacillus agaradhaerens* (*17*)] and fungi [e.g., *Paecilomyces varioti* (*18*, *19*), *Thermomyces lanuginosus* (*20*), *Trichoderma reesei* (*21*−*23*), *Aspergillus niger* (*24*), and *Aspergillus kawachii* (*18*)].

Structures have been reported for family 10 xylanases from both bacteria [e.g., *Cellulomonas fimi* (*25*−*29*), *Pseudomonas fluorescens* (*30*, *31*), *Streptomyces lividans* (*32*, *33*), *Streptomyces olivaceoviridis* (*34*, *35*)] and fungi [e.g., *Penicillium simplicissimum* (*36*, *37*) and *Thermoascus aurantiacus*

[†] This work was supported by a grant from the National Institutes of Health and contracts from the National Aeronautics and Space Administration.

[‡] PDB entry 1NOF.

* To whom correspondence should be addressed: Department of Molecular Biology and Biochemistry, 560 Steinhaus Hall, University of California, Irvine, CA 92697-3900. Telephone: (949) 824-1931. Fax: (949) 824-1954. E-mail: amcphers@uci.edu.

[§] Department of Molecular Biology and Biochemistry, University of California, Irvine.

[||] Current address: Molecular Biology Department, Instituto Potosino de Investigación Científica y Tecnológica, San Luis Potosí, SLP, México.

[⊥] Department of Plant Pathology, University of California, Riverside.

[@] Deceased.

[1] Abbreviations: GH, glycoside hydrolase; GH-5, glycoside hydrolase family 5; GH-10, glycoside hydrolase family 10; GH-30, glycoside hydrolase family 30; CD, catalytic module; CBM, carbohydrate-binding module; XBM, small module or putative xylan-binding module; XynA, xylanase from *E. chrysanthemi*; MIR, multiple isomorphous replacement; rms, root-mean-square.

10.1021/bi034144c CCC: $25.00 © 2003 American Chemical Society
Published on Web 06/26/2003

(*38*−*40*)]. These enzymes are considerably larger, at more than 400 amino acids in most cases, and generally contain a catalytic domain or module (CD) and a smaller carbohydrate-binding module (CBM). The xylanase from *P. simplicissimum* lacks a CBM. Of the GH-10 structures reported to date, all but one (*S. olivaceoviridis*) contain only the isolated catalytic domains. Although the xylanase from *S. olivaceoviridis* was crystallized with both the CD and the CBM, the connecting Gly/Pro-rich linker between the two modules was not evident in electron density maps. which suggests the possibility that the CBM is loosely attached to the CD and that the position of the CBM in the crystal structure may be determined more by packing than by interactions with the catalytic domain.

As previously reported by Keen et al. (*41*), xylanase from *Erwinia chrysanthemi* (designated XynA in this paper) is significantly homologous to vertebrate cerebrosidases, which are members of family 30, and to family 5 glycoside hydrolases. XynA has since been classified in GH-5 (*1*). A BLAST search of the PIR-NREF database at the PIR website (*42*), utilizing the full 413-amino acid sequence of XynA, found 13 GH-5 sequences with levels of identity in the range of 25−82% and levels of homology in the range of 44−90% and 18 GH-30 sequences with levels of identity in the range of 20−27% and levels of homology in the range of 36−45%. The endoxylanase from *Pectobacterium chrysanthemi* is 82% identical to XynA. All 13 GH-5 proteins possessing some level of sequence identity with XynA have been identified as having xylanase activity. No structures of these 13 enzymes have been determined to date. GH-5 enzymes contain seven invariant residues in five conserved stretches corresponding to five of the eight $\beta$-strands of the $\beta$-barrel. Four of the five stretches containing six of the seven invariant residues are found in XynA. On the basis of this sequence similarity, the catalytic residues were postulated to be Glu165 and Glu253 (*41*). These structural studies have confirmed these assignments.

XynA was first isolated from cultures of *E. chrysanthemi* SR120A, obtained from corn, and characterized as having xylanase activity by Braun and Rodrigues (*43*). Subsequently, Keen et al. (*41*) cloned the gene in *Escherichia coli* and obtained the secreted protein from the bacterial periplasm. Recently, XynA was found to be most active on xylan substrates substituted with α-1,2-linked 4-*O*-methyl-D-glucuronic acid, cleaving the substrate between the first and second xylose units from the glucuronic acid-substituted xylose on the nonreducing end (*44*). Despite these studies, no experimental determinations of the catalytic mechanism, the catalytic residues, or the function of the small domain have been reported. In the absence of such determinations, we can only speculate about these issues by analogy to homologous structures. Thus, since GH-5 enzymes operate through a configuration-retaining mechanism with glutamic acids on $\beta$-strands 4 and 7 acting as catalytic residues (*1*−*9*), XynA is expected to function by the same mechanism.

## EXPERIMENTAL PROCEDURES

*Production, Preparation, and Crystallization of the Enzyme.* Xylanase from *E. chrysanthemi* was cloned and expressed in *E. coli* and purified as previously described (*41*). As reported elsewhere, the purified enzyme crystallized reproducibly by vapor diffusion as mechanically sturdy parallelepiped-shaped crystals belonging to monoclinic space group $P2_1$ (No. 4) with the following cell dimensions: $a$ = 39.4 Å, $b$ = 49.6 Å, $c$ = 91.0 Å, $\beta$ = 101.7°, and $Z$ = 2 molecules/unit cell (*45*).

Table 1: Data Statistics[a]

| | |
|---|---|
| total no. of observations | 206521 |
| total no. of reflections | 58915 |
| resolution range (Å) | 1.42−44.57 (1.42−1.52) |
| redundancy | 3.51 (1.91) |
| $R_{sym}$ (%) | 6.56 (13.2) |
| $\langle I/\sigma(I)\rangle$ | 10.17 (3.14) |
| completeness (%) | 90.6 (61.9) |

[a] Data for the highest-resolution shell in parentheses.

*Data Collection.* Room-temperature diffraction data for both native and derivatized crystals were collected at a wavelength of 1.54 Å with a San Diego Multiwire Systems dual-detector system equipped with a Rigaku RU-200 rotating anode X-ray generator. Two crystals were used for the native data set. The native $R_{sym}$ was 0.066 for 206 521 observations, yielding a unique set of native data containing 58 915 reflections with a redundancy of 3.5 observations per reflection (see Table 1).

*Structure Solution.* Attempts to solve the structure by molecular replacement using GH-10 xylanase models found in the Protein Data Bank (*46*) proved to be fruitless. Therefore, the focus was diverted to obtaining experimental phase information by multiple isomorphous replacement (MIR). Heavy atom-soaked crystals were prepared by adding to the crystallization drops containing native crystals a solution of a heavy atom compound. Crystals were mounted by conventional means in quartz capillaries and data collected at room temperature as for the native crystals described above. A total of 33 heavy atom derivative data sets were collected, each from a single crystal.

All 33 derivative data sets were input individually to the program SOLVE (*47*−*51*) which produced prospective heavy atom sites for 27 of the derivatives (resolution range of 3−11 Å). Analysis of the phasing power and $R_{cullis}$ from heavy atom site refinement and cross phasing of derivative difference maps reduced the number of useful derivatives to four. The heavy atom sites of these four derivatives were refined at resolution ranges of 7−11, 6−11, 5−11, 4.5−11, 4−11, 3.5−11, and 3−11 Å. On the basis of the phasing power and the refinement stability, the useful resolution ranges of these four data sets were identified as indicated in Table 2. Thus, 3−11 Å data were used for the $HgCl_2/AuCl_3$ derivative, 4−11 Å data for two other derivatives, and 7−11 Å data for the fourth derivative. The overall figure of merit for this final MIR phase set was 0.49.

Early in our efforts, the $HgCl_2/AuCl_3$ derivative was identified as an excellent single-site derivative producing a phasing power of >1.0 down to a resolution of 3 Å. Therefore, this derivative was used as a monitor for phase quality. Any phase set in question, whether derived from a model, a solvent-flattened MIR map [using the program PHASES (*52*)], or phase combination of MIR and model phases (also using the program PHASES), was used to phase a $HgCl_2/AuCl_3$ − native difference density map for 3−11 Å data. The peak height of the Hg/Au site (in $\sigma$ units) served as an indicator of the quality of the phase set. The final MIR

Table 2: Derivative Data and Phasing Statistics[a]

| | native | $HgCl_2$/$AuCl_3$ | $HgCl_2$ | $K_2PtCl_6Pt$ | benzoate |
|---|---|---|---|---|---|
| no. of reflections | 58915 | 8598 | 14387 | 15201 | 12575 |
| no. of reflections (3−11 Å) | 6841 | 6227 | 6063 | 6483 | 5533 |
| no. of common reflections | − | 6176 | 6019 | 6408 | 5471 |
| $R_{sym}$ | 0.066 | 0.074 | 0.071 | 0.050 | 0.061 |
| $R_{der-nat}$ | − | 0.134 | 0.174 | 0.060 | 0.109 |
| no. of sites | − | 1 | 1 | 2 | 2 |
| phasing resolution (Å) | − | 3.0−11.0 | 4.0−11.0 | 4.0−11.0 | 7.0−11.0 |
| $R_{Cullis}$ | − | 0.49 | 0.56 | 0.59 | 0.52 |
| centric phasing power | − | 1.34 | 1.04 | 0.83 | 1.74 |
| acentric phasing power | − | 1.53 | 1.20 | 1.17 | 1.33 |
| site 1 | | | | | |
| *Q* | − | 0.63 | 0.62 | 0.23 | 1.04 |
| *X* | − | 0.167 | 0.172 | 0.963 | 0.409 |
| *Y* | − | 0.000 | 0.005 | 0.232 | 0.515 |
| *Z* | − | 0.089 | 0.088 | 0.500 | 0.401 |
| *B* (Å$^2$) | − | 23.6 | 23.5 | 33.1 | 459.0 |
| nearest residue | − | C175 | C175 | M347 | H336 |
| site 2 | | | | | |
| *Q* | − | − | − | 0.25 | 0.26 |
| *X* | − | − | − | 0.243 | 0.236 |
| *Y* | − | − | − | 0.117 | 0.145 |
| *Z* | − | − | − | 0.462 | 0.460 |
| *B* (Å$^2$) | − | − | − | 191.8 | 2.0 |
| nearest residue | − | − | − | M360 | M360 |

[a] Data sets for 29 other derivatives were collected. For the resolution range of 3−11 Å, the $R_{der-nat}$ values ranged from 0.051 to 0.324. No sites were unequivocally identified for any of these derivatives.

phase set produced an Hg/Au peak of 67$\sigma$. The phase set obtained by solvent flattening of the MIR map produced a peak of 51$\sigma$. Phase extension to 2.5−20 Å produced a phase set that yielded a peak of only 34$\sigma$.

The model was built [using the programs O (*53*) for building and MAPMAN (*54*) for converting XPLOR (*55*−*57*) maps to O maps] according to the MIR, density-modified MIR, and phase-extended density-modified MIR maps, producing a model whose phase set produced a peak for the Hg/Au site of only 24$\sigma$. This served as a baseline for model comparison. The model phases were combined with the MIR phases, and new maps were generated at various resolutions and with solvent flattening and phase extension. The model was rebuilt to these maps, resulting in a phase set that produced an Hg/Au peak of 27$\sigma$ in the monitoring map. Attempted refinement of this model (using XPLOR) gave model phase sets that produced smaller peaks for the Hg/Au site. Phases from the best models (based on peak heights of the Hg/Au site) of successive rebuilds were combined with the MIR phases followed by solvent flattening (with PHASES) until the model was satisfactory for producing relatively unbiased model-phased $2F_o - F_c$ maps for model building. This occurred when the model alone resulted in a phase set that gave an Hg/Au peak of 29$\sigma$. It was also at this point that a small six-residue sequence (residues 286−291, YVWWYI) was identified in the maps from which the entire sequence could be fitted to the map with minimal subsequent modifications of the sequence assignment. The final model phase set produced an Hg/Au peak of 41.5$\sigma$ in the monitor difference map.

*Structure Refinement.* The structure was refined isotropically with XPLOR to an *R* of 0.149 ($R_{free}$ = 0.182; 1.42 Å < *d* < 50 Å and $F > 4\sigma_F$) with 271 water molecules in the model and a bulk solvent correction applied. Refinement was continued with the program SHELXL-97 (*58*) with all fully occupied non-hydrogen atoms treated anisotropically, partially occupied non-hydrogen atoms refined isotropically, and hydrogen atoms riding in ideal positions with *B* factors equal to 1.2 or 1.5 times the thermal parameter ($B/B_{eq}$) of the atom to which they were bound. Water molecules were automatically picked by the SHELWAT routine and then confirmed by difference density maps [$nF_o - (n-1)F_c$ maps] and hydrogen bonding considerations. For $F > 4\sigma_F$, the final *R* = 0.095 and $R_{free}$ = 0.151; for all data, *R* = 0.102, $R_{free}$ = 0.161, and *S* = 1.49 (see Table 3).

## RESULTS

*The Model.* The final model consists of 383 amino acids (residues 31−413 in the enzyme sequence), 522 water molecules, and one acetate ion. The rms deviations from geometrical ideality are as follows: 0.010 Å for bonds, 2.13° for angles, 17.7° for dihedral angles, 1.40° for improper angles, and 3.8° for the deviation in $\omega$ from planarity. The estimated average error in the coordinates is 0.06 Å, as determined by the method of Luzzati (*59*). Electron density maps around the segment of residues 198−203 containing the *cis*-Ala201 in $\beta$-strand 5 of the catalytic domain are shown in Figure 1. A Ramachandran plot (*61*), produced with the program PROCHECK (*62*), shows 90.4% of non-glycine and non-proline residues in the most favored regions, 8.7% in the additionally allowed regions, and three residues (Glu165, Asp257, and Trp289) in the generously allowed regions; there are no residues in the disallowed regions. Asp257 is in the coil region between $\beta$-strand 7 and

Table 3: Refinement Statistics

| resolution shell (Å) | no. of reflections | *R* factor (all data) | *R* factor ($F > 4\sigma_F$) | no. of reflections | $R_{free}$ (all data) | $R_{free}$ ($F > 4\sigma_F$) |
|---|---|---|---|---|---|---|
| ∞−3.06 | 5904 | 0.123 | 0.122 | 716 | 0.163 | 0.163 |
| 3.06−2.43 | 5737 | 0.097 | 0.095 | 619 | 0.154 | 0.151 |
| 2.43−2.12 | 5845 | 0.078 | 0.074 | 663 | 0.134 | 0.128 |
| 2.12−1.93 | 5828 | 0.092 | 0.082 | 655 | 0.150 | 0.133 |
| 1.93−1.79 | 5519 | 0.078 | 0.073 | 618 | 0.155 | 0.147 |
| 1.79−1.68 | 5415 | 0.085 | 0.076 | 587 | 0.166 | 0.154 |
| 1.68−1.60 | 5302 | 0.099 | 0.083 | 606 | 0.173 | 0.150 |
| 1.60−1.53 | 5222 | 0.111 | 0.088 | 588 | 0.186 | 0.162 |
| 1.53−1.47 | 5123 | 0.130 | 0.098 | 580 | 0.202 | 0.171 |
| 1.47−1.4207 | 3048 | 0.146 | 0.100 | 340 | 0.211 | 0.184 |
| ∞−1.42 | 52943 | 0.102 | 0.095 | 5972 | 0.161 | 0.151 |
| ∞−1.42 | 58915 | 0.108 | 0.100 | | | |

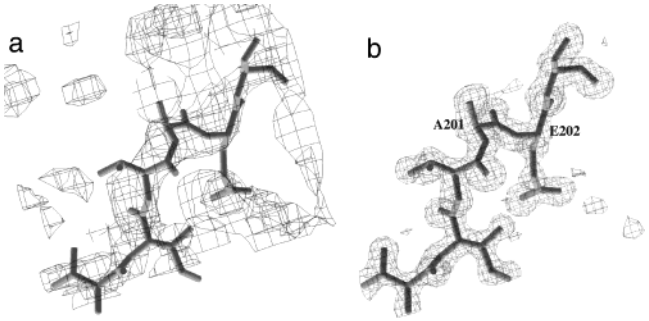


FIGURE 1: (a) MIR map (3 Å) contoured at 1.3$\sigma$. (b) Powell-minimized omit map (1.42 Å) with the displayed segment omitted and contoured at 3.1$\sigma$. The density clearly shows Ala201 is a *cis* peptide. This figure was produced with the program SETOR (*60*).

Table 4: Discretely Disordered Residues

| residue | position | residue | position |
|---|---|---|---|
| Asn40 | XBM $\beta_1$ | Lys90 | CD $\alpha_2$ |
| Lys107 | CD $\beta_3$ | Asn127 | CD $\beta_3\alpha_3$-loop |
| Ser137 | CD $\alpha_3$ | Ser141 | CD $\alpha_3$ |
| Asp145 | CD $\alpha_3$ | Ser147 | CD $\alpha_3$ |
| Thr152 | CD $\alpha_3$ | Asp171 | CD $\beta_4\alpha_4$-loop |
| Ser184 | CD $\alpha_4$ | Ser195 | CD $\alpha_4\beta_5$-loop |
| Lys197 | CD $\beta_5$ | Leu204 | CD $\beta_5\alpha_5$-loop |
| Lys216 | CD $\alpha_5$ | Ser218 | CD $\alpha_5$ |
| Ser221 | CD $\alpha_5$ | Val256 | CD $\beta_7\alpha_7$-loop |
| Lys259 | CD $\beta_7\alpha_7$-loop | Asn264 | CD $\beta_7\alpha_7$-loop |
| Lys303 | CD $\beta_8\alpha_8$-loop | Gln326 | XBM $\beta_2$ |
| Ser333 | XBM $\beta_2\beta_3$-loop | Asn334 | XBM $\beta_2\beta_3$-loop |
| Asn364 | XBM $\beta_5$ | Ser389 | XBM $\beta_7$ |
| Ser395 | XBM $\beta_7\beta_8$-loop | Val411 | XBM $\beta_9$ |

Table 5: Temperature Factor Analysis for the Final Model and Distribution of Water Molecules (3025 hydrogen atoms not included)

| atoms | count | min *B* | max *B* | mean *B* |
|---|---|---|---|---|
| protein | | | | |
| all protein atoms | 3047 | 2.8 | 97.8 | 13.6 |
| all main chain atoms | 1149 | 2.8 | 46.0 | 10.2 |
| all backbone atoms [poly(Ala)] | 1909 | 2.8 | 60.9 | 11.2 |
| all side chain atoms (with CB and O) | 1898 | 3.5 | 97.8 | 15.6 |
| all side chain atoms (no CB or O) | 1138 | 3.5 | 97.8 | 17.5 |
| water | | | | |
| all water O atoms | 522 | 6.2 | 69.8 | 36.5 |
| first-shell water (all) | 430 | 6.2 | 69.8 | 35.1 |
| first-shell water ($Q = 1.0$) | 291 | 6.2 | 69.8 | 36.4 |
| first-shell water ($Q = 0.5$) | 48 | 26.7 | 63.1 | 40.3 |
| first-shell water (disordered) | 78 | 8.0 | 54.1 | 27.2 |
| first-shell water (disordered) | 13 | 20.7 | 50.3 | 34.7 |
| second-shell water (all) | 92 | 26.2 | 65.7 | 43.0 |
| second-shell water ($Q = 1.0$) | 28 | 27.1 | 65.7 | 47.0 |
| second shell water ($Q = 0.5$) | 60 | 26.2 | 63.2 | 41.5 |
| second-shell water (disordered) | 4 | 29.6 | 47.6 | 38.6 |
| acetate | | | | |
| acetate ion | 4 | 13.6 | 28.9 | 23.4 |

α-helix 7. Glu165 and Trp289 are conserved residues in GH-5 enzymes. The first is the catalytic acid or base, and the second is positioned near the end of $\beta$-strand 8 spatially adjacent to the catalytic nucleophile, Glu253. There are 28 disordered residues that have been modeled with multiple conformations, 21 in the catalytic domain and seven in the small domain. These are listed in Table 4. The mean $B/B_{eq}$ values for the main chain and side chain atoms are 10 and 16 Å$^2$, respectively. Thermal parameter statistics are given in Table 5. In Figure 2, the enzyme is colored according to $B/B_{eq}$ value, with blue corresponding to the lowest and red to white corresponding to the highest values of $B/B_{eq}$. The water molecules and acetate ion have been included.

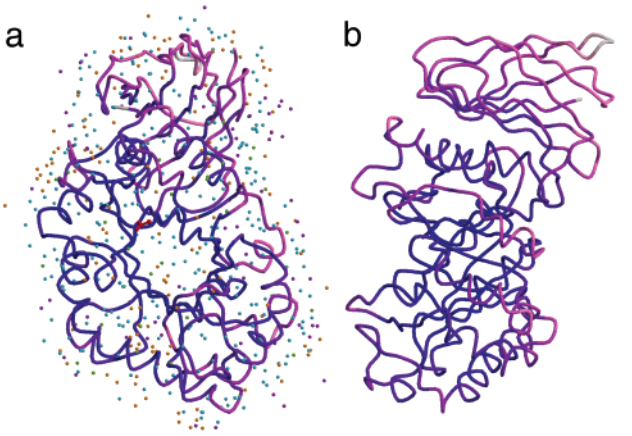


FIGURE 2: (a) Ribbon diagram of xylanase from *E. chrysanthemi* (XynA) colored by *B* factor (low *B* is blue and high *B* red to white) showing the toroidal shape of the catalytic domain (CD), the secondary structure, and the distribution of solvent. The acetate ion is red. The water molecules are color coded according to hydration shell and occupancy: cyan for first shell and full occupancy, orange for first shell and partial occupancy, green for second shell and full occupancy, and purple for second shell and partial occupancy. (b) Side view of XynA illustrating the disposition of the XBM toward the back side of the catalytic domain and the interface between α-helices 7 and 8 of the CD and the $\beta$-sheet of the XBM. This figure was rendered with the programs MOLSCRIPT (*63*) and RASTER3D (*64*).

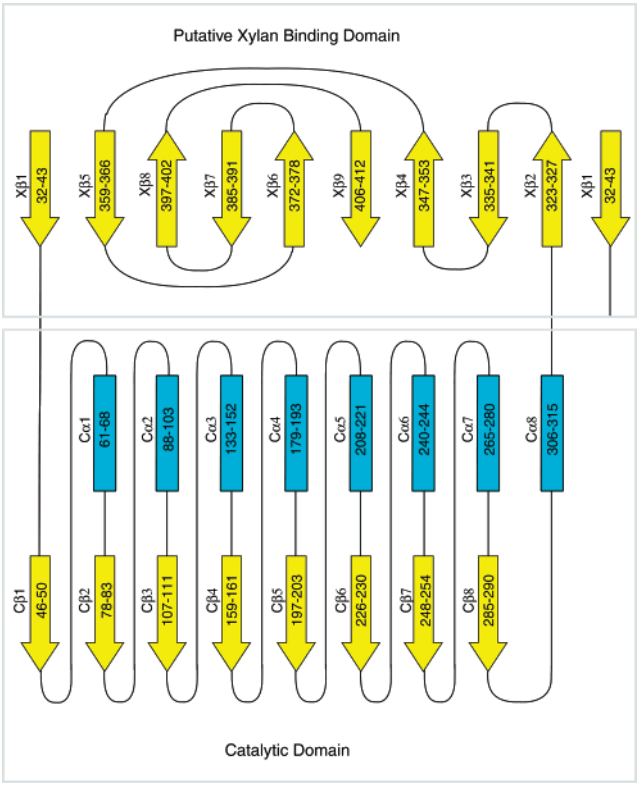


FIGURE 3: Schematic topological diagram of the secondary structure illustrating the involvement in the $\beta$-sandwich of the XBM of the first 10 residues of the N-terminus. This figure was produced with Adobe Illustrator 8.0 (*65*).

*Overall Structure.* A topological diagram of the xylanase secondary structure is shown in Figure 3. XynA is composed of a catalytic domain (CD) and a small domain with two linker peptides connecting them. The function of the small

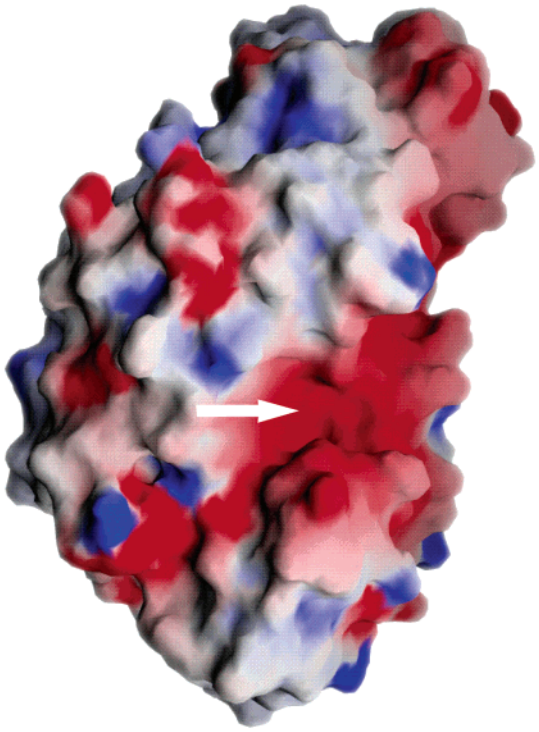

FIGURE 4: Illustration of the electrostatic potential of the enzyme surface with negatively charged regions colored red and positively charged regions colored blue. The large red region near the middle of the catalytic domain, marked by the arrow, is the catalytic binding cleft in which the negatively charged catalytic residues reside. This figure was generated with the program GRASP (*66*).

domain has not been determined, but by analogy to GH-10 xylanases, it is probably a carbohydrate-binding, or more specifically, a xylan-binding module. In this paper, it will be termed the small domain or module but abbreviated as XBM. The first 30 residues (the leader peptide) are missing from the structure due to cleavage by *E. coli* during secretion to the bacterial periplasm (*41*). Figure 2 illustrates the secondary structure and the molecular shape. The enzyme is approximately 68 Å long by 45 Å wide by 38 Å thick.

*Catalytic Domain.* The catalytic domain has the common $(\beta/\alpha)_8$-barrel folding motif consisting of residues 46−315. The $\beta$-barrel is elliptical in shape, and the connecting loops to the α-helices are generally shorter on the N-terminal side of the barrel (∼4.5 residues/loop) than are the loops on the C-terminal side (∼11.2 residues/loop; subsequently termed the back side and front side of the barrel, respectively). As is generally the case in clan GH-A glycoside hydrolases, the longer loops of the C-terminal side of the $\beta$-barrel take part in the formation of the binding cleft around the active site.

*Catalytic Binding Cleft.* The active site residues, Glu165 and Glu253, are positioned on opposite sides of the catalytic binding cleft at the C-terminal ends of $\beta$-strands 4 and 7. Glu165 is actually situated in the loop following $\beta$-strand 4 and, by analogy to related structures, serves as the acid or base. Glu253 is the nucleophile during catalytic hydrolysis. The carboxyl oxygen atoms of closest approach of the catalytic residues are 4.22 Å apart. The cleft is decidedly negative as shown in Figure 4. However, the binding cleft contains a number of aromatic residues that can serve as the scaffolding for the docking of the pyranose rings of the substrate. These include Trp55, Tyr172, Trp168, and Tyr232. Trp113, Trp289, and Tyr290 are buried and accessible only as possible hydrogen-bonding donors.

*Small Domain.* The XBM has nine $\beta$-strands forming an elliptical $\beta_9$-barrel. The first strand is composed of residues 31−43, the first 13 residues of the secreted enzyme. The last 91 residues in the sequence form the other eight $\beta$-strands (see Figure 3). Both linker peptides connect the XBM to the CD on the back side of the catalytic domain. As seen in Figure 2b, the XBM is positioned somewhat to the back side of the CD. Strands 2−4, 6, and 9 of the small module form the interface with α-helices 7 and 8 of the CD. The interface is distributed between hydrophobic and hydrophilic residues and includes 11 hydrogen bonds between α-helices 7 and 8 and the adjacent $\beta$-sheet. Approximately 1200 Å$^2$ of each domain is buried by the formation of this interface with only 10% contributed by charged residues. Interactions involved in this interface are listed in Table 6.

*Water Molecules and the Acetate Ion.* The water structure is quite extensive. There are 319 water molecules with full occupancy, 108 water molecules with 50% occupancy, and 95 disordered water molecules (defined as sharing space with other atoms in the model). There are 430 waters in the first hydration shell and 92 waters in the second hydration shell of the enzyme. The distribution of modeled waters is listed in Table 5.

The acetate ion, which has an average $B_{eq}$ of 23 Å$^2$, is bound to Arg293, bridging atoms $N_\epsilon$ and $N_\zeta 1$ while also interacting with the OH group of Tyr295.

*Molecular Packing.* Xylanase molecules have 10 neighbors in the unit cell. Table 7 lists the hydrogen-bonded interactions between the reference molecule and its neighbors. This list includes salt bridges between Glu176 and Arg99 and between Glu181 and Arg99. The XBM in XynA has possibly eight stabilizing hydrogen bonds with neighboring molecules, all but two of these being formed with XBMs of neighboring molecules. There are 41 water molecules involved in intermolecular interactions (within 3.5 Å of two xylanase molecules).

## DISCUSSION

*Comparison to Family 5 and Family 10 Glycoside Hydrolases.* Comparison of XynA to the known GH-5 and GH-10 structures identifies Glu165 and Glu253 as the catalytic residues. This confirms the originally proposed assignments (*41*). Unlike many GH-5 and GH-10 enzymes, XynA contains no disulfide bonds. Most disulfide bonds serve to stabilize floppy loops, but the shorter loops of XynA apparently need no such stabilization.

Using the Smith−Waterman algorithm (*42*), pairwise alignment of the XynA sequence with sequences of family 5 glycoside hydrolases found two xylanases of *Aeromonas caviae* to be more than 40% identical for 275 residues of the XynA catalytic domain sequence. Attempts to align the XynA CD sequence (275 amino acids) with a number of other family 5 sequences and with family 10 xylanases never resulted in a simultaneous match of both catalytic residues. Clearly, the fact that GH-5 and GH-10 families are members of the GH-A clan implies that they share some common features, namely, a common fold, the same type of catalytic mechanism, and several conserved residues (*11*). Since the level of sequence similarity is low, we compare XynA with

Table 6: (A) Accessible Surface Area [calculated with the algorithm of Lee and Richards (*67*) as implemented in XPLOR (*55*−*57*)] of Xylanase XynA and Its Domains and Buried Surface Area in the Catalytic Domain−Xylan Binding Module Interface Showing that the Interface Is Dominated by Hydrophobic and Neutral Hydrophilic Residues and (B) Possible Hydrogen Bonding Interactions between the Catalytic Domain and the Small Module

(A)

| domain | total area | hydrophobic | aromatic | neutral hydrophilic | charged hydrophilic |
|---|---|---|---|---|---|
| XynA | 14229 | 3146 | 1414 | 5266 | 4402 |
| % by residue type | | 22 | 10 | 37 | 31 |
| CD (in the presence of XBM) | 9645 | 2360 | 1016 | 3174 | 3095 |
| % of total ASA of XynA | 68 | 75 | 72 | 60 | 70 |
| XBM (in the presence of CD) | 4584 | 786 | 398 | 2093 | 1308 |
| % of total ASA of XynA | 32 | 25 | 28 | 40 | 30 |
| CD (in the absence of XBM) | 10905 | 2829 | 1362 | 3509 | 3206 |
| % of total CD ASA | | 26 | 12 | 32 | 29 |
| XBM (in the absence of CD) | 5777 | 1336 | 496 | 2519 | 1426 |
| % of total XBM ASA | | 23 | 9 | 44 | 25 |
| buried surface area | | | | | |
| CD | 1260 | 469 | 346 | 335 | 111 |
| % of total buried ASA of CD | | 37 | 27 | 27 | 9 |
| XBM | 1193 | 550 | 98 | 426 | 118 |
| % of total buried ASA of XBM | | 46 | 8 | 36 | 10 |

(B)

| CD residue | | XBM residue | distance (Å) | location |
|---|---|---|---|---|
| CD $\alpha_7$ | | | | |
| 265 NE1 | → | 405 O | 2.85 | $\beta_8\beta_9$-loop |
| 276 ND2 | → | 329 OE1 | 2.94 | $\beta_2\beta_3$-loop |
| 279 O | ← | 324 NE | 3.01 | XBM $\beta_2$ |
| CD $\alpha_8$ | | | | |
| 306 NZ | → | 379 O | 2.85 | $\beta_6\beta_7$-loop |
| 306 NZ | → | 380 O | 3.38 | $\beta_6\beta_7$-loop |
| 306 NZ | → | 382 O | 3.24 | $\beta_6\beta_7$-loop |
| 306 NZ | → | 383 OD1 | 2.74 | $\beta_6\beta_7$-loop |
| 313 OE1 | ← | 378 N | 3.10 | XBM $\beta_6$ |
| 313 OE1 | ← | 378 OG1 | 2.70 | XBM $\beta_6$ |
| 313 NE2 | → | 376 O | 2.94 | XBM $\beta_6$ |
| 313 NE2 | → | 409 OG1 | 2.89 | XBM $\beta_9$ |
| 313 O | ← | 409 OG1 | 2.80 | XBM $\beta_9$ |

Table 7: Intermolecular Hydrogen Bonding in Xylanase XynA

| reference | symop[a] | bonded atom | distance (Å) | reference | symop[a] | bonded atom | distance (Å) |
|---|---|---|---|---|---|---|---|
| Thr32 OG1 | 7 | Asn244 OD1 | 2.77 | Trp177 O | 2 | Ser103 OG | 3.28 |
| Lys34 NZ | 7 | Asn244 O | 2.95 | Glu181 OE1 | 2 | Arg99 NH2 | 2.86 |
| Asn38 OD1 | 10 | Ser389 N | 2.87 | Glu181 OE2 | 2 | Arg99 NH1 | 2.76 |
| | 10 | Ser389 OG A[b] | 3.09 | | 2 | Arg99 NH2 | 3.48 |
| Asn57 ND2 | 5 | Tyr118 OH | 3.27 | Ser191 OG | 4 | Gln260 NE2 | 2.89 |
| Asn65 ND2 | 1 | Asp167 O | 3.37 | Ser191 O | 4 | Asn264 ND2 A[b] | 2.96 |
| Ser88 OG | 6 | Arg293 NH1 | 2.99 | | 4 | Asn264 ND2 B[b] | 3.10 |
| | 6 | Arg293 NH2 | 3.00 | Asn244 OD1 | 8 | Thr32 OG1 | 2.77 |
| Ser89 OG | 6 | Ser294 OG | 2.98 | Asn244 O | 8 | Lys34 NZ | 2.95 |
| Pro96 O | 1 | Lys169 NZ | 3.28 | Gln260 NE2 | 3 | Ser191 OG | 2.89 |
| Arg99 NE | 1 | Glu176 OE1 | 2.82 | Asn264 ND2 A[b] | 3 | Ser191 O | 2.96 |
| Arg99 NH1 | 1 | Glu181 OE2 | 2.76 | Arg292 NH1 | 5 | Tyr118 OH | 3.20 |
| Arg99 NH2 | 1 | Glu181 OE1 | 2.86 | Arg293 NH1 | 5 | Ser88 OG | 2.99 |
| | 1 | Glu181 OE2 | 3.48 | Arg293 NH2 | 5 | Ser88 OG | 3.00 |
| Gln100 NE2 | 1 | Asp167 O | 2.87 | Ser294 OG | 5 | Ser89 OG | 2.98 |
| Ser103 OG | 1 | Trp177 O | 3.28 | Glu300 OE1 | 5 | Tyr118 OH | 3.12 |
| Tyr118 OH | 6 | Asn57 ND2 | 3.27 | Thr343 O | 10 | Ser399 OG1 | 2.85 |
| | 6 | Arg292 NH1 | 3.20 | Asp358 OD2 | 9 | Asn369 ND2 | 2.69 |
| | 6 | Glu300 OE1 | 3.12 | Asn369 ND2 | 10 | Asp358 OD2 | 2.69 |
| Arg130 NH1 | 2 | Thr152 O | 3.20 | | 10 | Asn403 OD1 | 3.03 |
| Thr152 O | 1 | Arg130 NH1 | 3.20 | Ser389 N | 9 | Asn38 OD1 | 2.87 |
| Asp167 O | 2 | Asn65 ND2 | 3.37 | Ser389 OG A[b] | 9 | Asn38 OD1 | 3.09 |
| | 2 | Gln100 NE2 | 2.87 | Thr399 OG1 | 9 | Thr343 O | 2.85 |
| Lys169 NZ | 2 | Pro96 O | 3.28 | Asn403 OD1 | 9 | Asn369 ND2 | 3.03 |
| Glu176 OE1 | 2 | Arg99 NE | 2.82 | Asn264 ND2 B[b] | 3 | Ser191 O | 3.10 |

[a] Symmetry operators: $x + 1, y, z$ (1); $x - 1, y, z$ (2); $x, y + 1, z$ (3); $x, y - 1, z$ (4); $1 - x, y + {}^1/_2, -z$ (5); $1 - x, y - {}^1/_2, -z$ (6); $1 - x, y + {}^1/_2, 1 - z$ (7); $1 - x, y - {}^1/_2, 1 - z$ (8); $2 - x, y + {}^1/_2, 1 - z$ (9); and $2 - x, y - {}^1/_2, 1 - z$ (10). [b] Alternate conformations.

GH-5 and GH-10 enzymes by a structural alignment with those structures that have been determined crystallographically.

A simple structural alignment of XynA with models of GH-5 and GH-10 enzymes is summarized in Table 8. Coordinates were obtained from the Protein Data Bank (*46*).

Table 8: Structural Alignment of Xylanase with Other Family 5 Glycoside Hydrolases and Family 10 Xylanases[a]

| | | 1A3H | 1BQC | 1CEC | 1CZ1 | 1ECE | 1EDG | 1EGZ | 1QNP | 1B31 | 1CLX | 1T*X | 1XYF | 1XYZ | 2EXO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| no. of residues | | 300 | 302 | 331 | 394 | 358 | 380 | 291 | 344 | 301 | 345 | 302 | 427 | 320 | 312 |
| family | | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 10 | 10 | 10 | 10 | 10 | 10 |
| XynA | eq | 114 | 114 | 99 | 125 | 122 | 119 | 112 | 124 | 123 | 110 | 120 | 123 | 119 | 123 |
| | id | 19 | 20 | 10 | 18 | 21 | 11 | 19 | 17 | 21 | 14 | 19 | 18 | 12 | 20 |
| | rmsd | 1.33 | 1.28 | 1.42 | 1.27 | 1.26 | 1.19 | 1.34 | 1.34 | 1.55 | 1.47 | 1.51 | 1.41 | 1.46 | 1.41 |
| 1A3H | eq | | 168 | 154 | 142 | 183 | 153 | 271 | 132 | 120 | 112 | 115 | 110 | 100 | 116 |
| | id | | 31 | 19 | 32 | 32 | 25 | 117 | 21 | 13 | 18 | 10 | 9 | 8 | 12 |
| | rmsd | | 1.22 | 1.30 | 1.32 | 1.27 | 1.21 | 0.81 | 1.33 | 1.43 | 1.42 | 1.30 | 1.39 | 1.41 | 1.49 |
| 1BQC | eq | | | 159 | 141 | 170 | 146 | 159 | 159 | 105 | 119 | 120 | 106 | 102 | 96 |
| | id | | | 31 | 28 | 41 | 35 | 32 | 28 | 10 | 18 | 17 | 13 | 10 | 12 |
| | rmsd | | | 1.17 | 1.21 | 1.26 | 1.33 | 1.38 | 1.28 | 1.43 | 1.31 | 1.31 | 1.34 | 1.40 | 1.45 |
| 1CEC | eq | | | | 203 | 183 | 224 | 153 | 167 | 86 | 103 | 98 | 97 | 85 | 88 |
| | id | | | | 46 | 38 | 47 | 29 | 29 | 11 | 12 | 9 | 13 | 8 | 10 |
| | rmsd | | | | 1.10 | 1.16 | 1.12 | 1.19 | 1.22 | 1.41 | 1.40 | 1.41 | 1.32 | 1.44 | 1.42 |
| 1CZ1 | eq | | | | | 171 | 187 | 152 | 166 | 110 | 125 | 119 | 115 | 128 | 120 |
| | id | | | | | 35 | 42 | 27 | 39 | 14 | 16 | 14 | 13 | 14 | 14 |
| | rmsd | | | | | 1.23 | 1.19 | 1.38 | 1.29 | 1.37 | 1.41 | 1.37 | 1.41 | 1.35 | 1.34 |
| 1ECE | eq | | | | | | 176 | 176 | 182 | 117 | 144 | 127 | 124 | 126 | 122 |
| | id | | | | | | 42 | 40 | 41 | 23 | 21 | 21 | 15 | 20 | 17 |
| | rmsd | | | | | | 1.23 | 1.31 | 1.30 | 1.36 | 1.39 | 1.45 | 1.41 | 1.37 | 1.40 |
| 1EDG | eq | | | | | | | 155 | 166 | 123 | 129 | 121 | 121 | 100 | 124 |
| | id | | | | | | | 30 | 28 | 22 | 15 | 18 | 16 | 14 | 16 |
| | rmsd | | | | | | | 1.30 | 1.41 | 1.36 | 1.38 | 1.38 | 1.42 | 1.46 | 1.40 |
| 1EGZ | eq | | | | | | | | 128 | 117 | 119 | 122 | 120 | 121 | 118 |
| | id | | | | | | | | 29 | 15 | 17 | 16 | 18 | 17 | 15 |
| | rmsd | | | | | | | | 1.42 | 1.42 | 1.39 | 1.33 | 1.42 | 1.32 | 1.34 |
| 1QNP | eq | | | | | | | | | 101 | 122 | 104 | 132 | 111 | 100 |
| | id | | | | | | | | | 17 | 11 | 14 | 20 | 17 | 14 |
| | rmsd | | | | | | | | | 1.47 | 1.43 | 1.37 | 1.35 | 1.43 | 1.42 |
| 1B31 | eq | | | | | | | | | | 244 | 300 | 268 | 254 | 263 |
| | id | | | | | | | | | | 98 | 219 | 129 | 112 | 107 |
| | rmsd | | | | | | | | | | 0.99 | 0.62 | 0.92 | 0.98 | 1.01 |
| 1CLX | eq | | | | | | | | | | | 247 | 256 | 252 | 250 |
| | id | | | | | | | | | | | 112 | 117 | 96 | 115 |
| | rmsd | | | | | | | | | | | 0.98 | 1.01 | 1.08 | 1.05 |
| 1T*X | eq | | | | | | | | | | | | 272 | 260 | 263 |
| | id | | | | | | | | | | | | 138 | 118 | 119 |
| | rmsd | | | | | | | | | | | | 0.87 | 1.03 | 0.91 |
| 1XYF | eq | | | | | | | | | | | | | 277 | 287 |
| | id | | | | | | | | | | | | | 122 | 153 |
| | rmsd | | | | | | | | | | | | | 0.92 | 0.93 |
| 1XYZ | eq | | | | | | | | | | | | | | 292 |
| | id | | | | | | | | | | | | | | 127 |
| | rmsd | | | | | | | | | | | | | | 1.01 |

Average Values of Spatially Equivalent Residues, Identical Residues, and rms Deviations for Various Structural Groups of Glycoside Hydrolases (standard deviations are given)

| | XynA−family 5 | XynA−family 10 | family 5−family 5 | family 5−family 10 | family 10−family 10 |
|---|---|---|---|---|---|
| eq | 116 ± 8 | 118 ± 6 | 169 ± 28 | 114 ± 12 | 266 ± 16 |
| id | 17 ± 2 | 16 ± 3 | 36 ± 17 | 15 ± 4 | 125 ± 29 |
| rmsd | 1.31 ± 0.07 | 1.47 ± 0.07 | 1.25 ± 0.12 | 1.40 ± 0.05 | 0.95 ± 0.11 |

[a] Except for XynA, the PDB entries for the various structures are used to identify the representative GH-5 and GH-10 glycoside hydrolases used in this analysis.

The alignment was determined by three-residue superpositioning followed by rigid body refinement using XPLOR in which the distances between $C_\alpha$ atoms of equivalent residues were minimized. During this process, "equivalent residues" were defined as those residues having $C_\alpha$ distances of $\leq 2$ Å. The subset of equivalent and identical residues and the rms deviation of the $C_\alpha$ atoms of the equivalent residues are listed. The averages of the individual superpositions, at the bottom of Table 8, reveal that xylanase from *E. chrysanthemi* is about as structurally similar (or not similar) to the family 5 glycoside hydrolases as it is to the family 10 xylanases. Its level of structural similarity to the family 10 xylanases is about the same as the average of the other GH-5 glycoside hydrolases to the GH-10 xylanases. However, in terms of the average number of equivalent (115 vs 169) and identical (17 vs 36) residues, xylanase XynA is less similar to the group of family 5 glycoside hydrolases used in this study than that group is to itself, although the average rms deviation between $C_\alpha$ atoms of equivalent residues is similar. In general, family 5 glycoside hydrolases are a more structurally diverse family than the family 10 xylanases as indicated by the average number of equivalent residues (169 vs 266) and identical residues (36 vs 125) and the average rms deviation of equivalent $C_\alpha$ atoms (1.25 Å vs 0.95 Å). This observed diversity has prompted a further classification of the GH-5 family into at least nine subfamilies (*68*). Figure 5 shows the superposition of a representative from each of the family 5 and family 10 glycoside hydrolase families onto XynA.

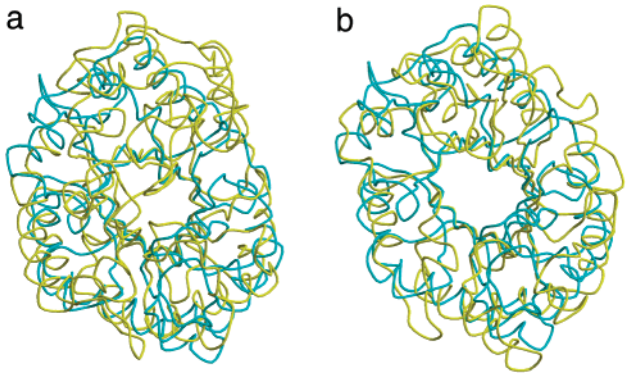


FIGURE 5: Superposition of catalytic domains of XynA (colored blue) onto (a) the GH-5 endoglucanase of *E. chrysanthemi* (PDB entry 1EGZ) and (b) the GH-10 xylanase of *P. simplicissimum* (PDB entry 1B31). The excellent fit of the XynA $\beta$-barrel to both is apparent; the differences in the front side loop structure and the α-helices are also obvious. All conserved residues are either in the $\beta$-strands of the barrel or near the beginning of the loops extending from the $\beta$-barrel strands. This figure was rendered with the programs MOLSCRIPT (*63*) and RASTER3D (*64*).

In each case, the $\beta$-barrels of the catalytic domains align well. There is little correlation in the positioning, orientation, and length of the helices surrounding the $\beta$-barrel; the wide variation in loop size, especially on the C-terminal side of the $\beta$-barrel, produces a broad spectrum of diverse loop structures.

*Catalytic Residues.* Generally, both the GH-5 and GH-10 families have a narrow range of $C_\alpha\cdots C_\alpha$ and $O_\epsilon\cdots O_\epsilon$ distances between catalytic residues in the uncomplexed structures. Except for the apparent anomaly of the endoglucanase of *Clostridium thermocellum* [PDB entry 1CEC (*69*)] in the GH-5 family, the $C_\alpha\cdots C_\alpha$ distances of the catalytic residues in this family range from 9.67 to 10.77 Å and the $O_\epsilon\cdots O_\epsilon$ distances range from 3.13 to 4.20 Å. In XynA, these quantities are near this range: 10.01 and 4.22 Å, respectively. The *C. thermocellum*, on the other hand, has distances of 12.74 and 10.05 Å for the $C_\alpha\cdots C_\alpha$ and $O_\epsilon\cdots O_\epsilon$ distances, respectively. However, once the substrate binds, the loop containing Glu140 (the catalytic acid or base) moves 3 Å with rotation of the Glu140 side chain by ~90°, resulting in an $O_\epsilon\cdots O_\epsilon$ distance of 4.3 Å, consistent with the other GH-5 enzymes (*70*). The corresponding ranges for the GH-10 xylanases are 12.50−12.72 and 5.08−5.62 Å, respectively, which are very distinct from the GH-5 ranges.

*Conserved Residues.* Perusal of the literature reveals some confusion concerning the conserved residues in the GH-5 glycoside hydrolases. Some references state that there are eight conserved residues (*71*−*73*), while others claim that there are seven (*74*, *75*). The seven consensus residues are (1) an arginine in $\beta$-strand 2, (2) a histidine in the $\beta_3\alpha_3$-loop, (3 and 4) the asparagine-glutamic acid catalytic pair at the end of $\beta$-strand 4, (5 and 6) a His-X-Tyr tripeptide at the end of $\beta$-strand 6, and (7) the other catalytic glutamic acid at the end of $\beta$-strand 7. In those papers that claim there are eight conserved residues, the eighth residue is a tryptophan in $\beta$-strand 8. Table 9 demonstrates that there is good alignment of the conserved residues among the known family 5 glycoside hydrolases. However, it is clear that XynA has a tryptophan in the place of the histidine in the $\beta_3\alpha_3$-loop. The sequence alignment between XynA and nine family 5 glycoside hydrolases that are more than 25% identical and

Table 9: Conserved Residues and Distances between $C_\alpha$ Positions in the Family 5 and Family 10 Glycoside Hydrolase Structures[a]

| structure | R | H | N | E | H | Y | E | W |
|---|---|---|---|---|---|---|---|---|
| | | | | Family 5 | | | | |
| XynA | R81 | W113 | N164 | E165 | H230 | Y232 | E253 | W289 |
| 1A3H | R62 | H101 | N138 | E139 | H200 | Y202 | E228 | W262 |
| | 0.15 | 1.93 | 1.28 | 0.71 | 0.69 | 3.61 | 0.63 | 1.05 |
| 1QNR | R54 | H102 | N168 | E169 | H241 | Y243 | E276 | W306 |
| | 0.47 | 24.75 | 1.76 | 1.58 | 0.61 | 2.38 | 0.69 | 1.08 |
| 1ECEA | R62 | H116 | N161 | E162 | H238 | Y240 | E282 | W319 |
| | 0.63 | 1.85 | 0.37 | 0.35 | 0.53 | 3.39 | 0.70 | 1.20 |
| 1CZ1 | R92 | H135 | N191 | E192 | H253 | Y255 | E292 | W363 |
| | 0.63 | 1.32 | 0.17 | 0.16 | 0.62 | 3.27 | 0.77 | 0.71 |
| 1EDG | R79 | H122 | N169 | E170 | H254 | Y256 | E307 | W340 |
| | 0.31 | 1.14 | 0.42 | 0.99 | 0.59 | 3.12 | 0.68 | 0.54 |
| 1CEC | R62 | H90 | N139 | E140 | H198 | Y200 | E280 | W313 |
| | 0.15 | 2.38 | 2.38 | 1.01 | 3.50 | 6.93 | 2.59 | 2.06 |
| 1EGZ | R57 | H98 | N132 | E133 | H192 | Y194 | E220 | W254 |
| | 0.42 | 1.95 | 1.16 | 0.74 | 0.41 | 4.05 | 0.53 | 0.65 |
| 1BQC | R50 | H86 | N127 | E128 | H196 | Y198 | E225 | W254 |
| | 1.08 | 0.68 | 0.15 | 0.47 | 0.58 | 2.94 | 0.53 | 0.35 |
| | | | | Family 10 | | | | |
| 1B31 | T45 | W88 | N131 | E132 | Q208 | H210 | E238 | W268 |
| | 0.72 | 3.71 | 2.37 | 2.50 | 1.50 | 2.79 | 0.22 | 1.31 |
| 2EXO | V41 | W84 | N126 | E127 | Q203 | H205 | E233 | W273 |
| | 0.61 | 3.90 | 2.81 | 2.76 | 1.11 | 2.50 | 0.31 | 1.45 |
| 1E0W | T42 | W85 | N127 | E128 | Q205 | H207 | E236 | W266 |
| | 1.41 | 3.38 | 1.51 | 1.41 | 1.60 | 3.79 | 1.41 | 2.10 |
| 1TAX | T44 | W87 | N130 | E131 | Q207 | H209 | E237 | W267 |
| | 0.72 | 4.05 | 2.92 | 2.79 | 1.54 | 3.07 | 1.58 | 2.80 |
| 1XYF | T42 | W85 | N127 | E128 | Q205 | H207 | E236 | W266 |
| | 0.30 | 4.63 | 2.75 | 2.53 | 1.06 | 2.62 | 0.49 | 1.58 |
| 1CLXA | T41 | W83 | N126 | E127 | Q213 | H215 | W246 | W305 |
| | 0.60 | 3.23 | 2.23 | 2.57 | 1.30 | 2.88 | 0.34 | 1.09 |
| 1XYZ | T557 | W600 | N644 | E645 | Q721 | H723 | E754 | W795 |
| | 0.83 | 4.49 | 2.65 | 2.69 | 1.03 | 2.02 | 0.23 | 1.26 |

[a] Except for XynA, the PDB entries for the various structures are used to identify the representative GH-5 and GH-10 glycoside hydrolases used in this analysis.

a tryptophan corresponding to Trp113 in the $\beta_3\alpha_3$-loop of XynA were all identified as xylanases. However, there are no three-dimensional structures of any of these enzymes to confirm the identical positioning of their tryptophan residues. One interesting anomaly to these correspondences is the $\beta$-mannanase from *T. reesei* [PDB entry 1QNR (*71*)]. Sequence alignment and hydrophobic cluster analysis identify His102 as the conserved histidine residue; there are no other histidine residues between position 102 and the catalytic acid or base, Glu169, in the sequence. However, the $C_\alpha$ atom of His102 is ~24 Å from the generally observed position of the corresponding conserved histidine residues. Instead, Asn112 lies near this structural position. Thus, although the sequence suggests His102 is conserved, structurally it is not conserved. Unfortunately, no structure of $\beta$-mannanase from *T. reesei* having substrate bound in site −1 has been determined. Therefore, the residue, if any, that plays the role of the conserved histidine (i.e., binding O3′ of the mannose at the −1 site) remains a mystery. This question arises: is the deposited sequence correct? If not, does the density at residue 112 in the 1.5 Å electron density maps suggest the presence of a histidine residue?

The putative conserved histidine in the GH-5 enzymes is replaced with a conserved tryptophan in the family 10 xylanases. Clearly, XynA and the other GH-5 xylanases mentioned above are similar to the GH-10 xylanases in this sense. This would suggest an important role in xylose

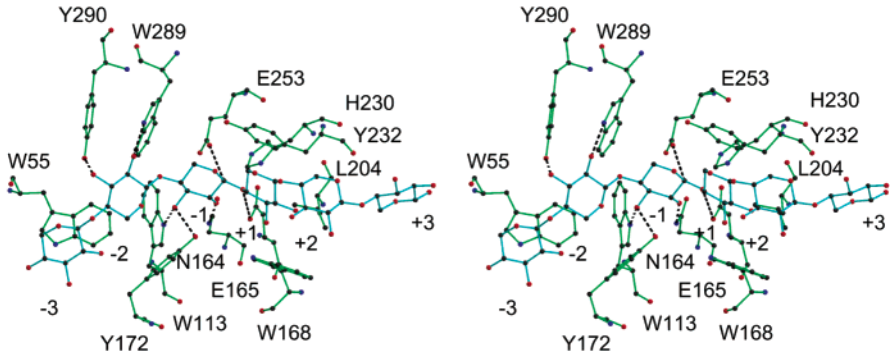


FIGURE 6: Stereoview of a postulated model of xylohexaose (blue) binding to XynA (green). Only the side chains of residues that would interact with the carbohydrate are included. The enzyme residues and all carbohydrate sites are labeled. Interactions between the protein and the xylohexaose are represented by black dashed lines as follows (protein to xylose): Y172 OH−O3 (−1 site), W113 $N_\epsilon$−O3 (−1 site), N164 $N_\delta$2−O2 (−1 site), E165 $O_\epsilon$2−O4 (+1 site), E253 $O_\epsilon$1−C1 (−1 site), W289 $N_\epsilon$1−O2 (−2 site), and Y290 OH−O3 (−2 site). The pyranose ring at the −3 site lies flat against the indole ring of Trp55. Leu204 sits in the channel formed by Trp168 and Tyr232, forcing the reducing end of the xylohexaose out of the channel and perhaps distorting the molecular structure at the −1 and +1 sites, thus facilitating bond breakage. This figure was rendered with the programs MOLSCRIPT (*63*) and RASTER3D (*64*).

Table 10: Results of a BLAST Search of the PIR-NREF Database Using the 94 C-Terminal Residues of XynA

| no. | organism | total no. of amino acids | CBM residues | CBM identity (%) | CD residues | CD identity (%) | identified activity[a] |
|---|---|---|---|---|---|---|---|
| 0 | *E. chrysanthemi* | 413 | 320−413 | 100 | 45−319 | 100 | XynA |
| 1 | *E. chrysanthemi* | 413 | 320−413 | 84 | 45−319 | 84 | xylanase |
| 2 | *X. campestris* | 405 | 313−405 | 44 | 39−312 | 62 | xylanase |
| 3 | *B. subtilis* | 422 | 330−419 | 39 | 48−329 | 41 | xylanase |
| 4 | *A. punctata* | 528 | 327−417 | 39 | 44−326 | 41 | xylanase |
| 5 | *A. punctata* | 564 | 332−422 | 38 | 49−331 | 41 | xylanase |
| 6 | *C. acetobutylicum* | 588 | 328−419 | 37 | 47−327 | 42 | xylanase |
| 7 | *C. acetobutylicum* | 586 | 327−417 | 35 | 48−324 | 42 | xylanase |
| 8 | *R. albus* | 726 | 336−427 | 31 | 53−335 | 40 | xylanase |
| 9 | *X. axonopodis* | 109 | 17−107 | 46 | − | − | truncated xylanase |

[a] Identified in the PIR-NREF database search.

specificity for this tryptophan. However, in the GH-10 enzymes, this tryptophan is hydrogen bonded to the catalytic acid residue, whereas no such bond is established in XynA. Instead, the tryptophan $N_\epsilon$1 superimposes nicely on the $N_\epsilon$2 atoms of the equivalent histidine residues of the GH-5 glycoside hydrolases.

*Active Site Residues and Substrate Binding.* Analysis of the structures of substrate- and/or inhibitor-bound glycoside hydrolases from both GH families 5 and 10 suggests a mode of binding of polyxylose to XynA, at least in nonreducing sites −1, −2, and −3. This postulated structure is illustrated in Figure 6. In the model, the xylose in site −1 interacts with Trp113, Asn164, Glu165, Tyr172, and Glu253; the xylose in site −2 is hydrogen bonded to Trp289 and Tyr290 and in contact with Tyr172, and the xylose ring in site −3 lies across the face of the indole ring of Trp55.

The path of the reducing end of a polyxylose substrate is difficult to discern. The aromatic rings of Trp168 and Tyr232 are parallel, forming a channel approximately 7.5 Å wide into which a polyxylose chain could be accommodated comfortably except for the presence of Leu204 in this channel. The presence of this leucine may assist in distorting the glycosidic linkage between the −1 and +1 moieties which would facilitate bond breakage. On the other hand, there is space into which the Leu204 side chain could rotate to allow more penetration of the polyxylose into the Trp/Tyr channel. According to Hurlbert and Preston (*44*), XynA has specificity for xylan containing an α-1,2-linked 4-*O*-methyl-D-glucuronic acid substituent attached to the xylose moiety in the +2 binding site. They state that the specificity is a result of carboxylate group recognition by the enzyme; how this recognition is achieved is unclear from modeling.

*Carbohydrate-Binding Modules.* Whether the apparently strong interface between the catalytic domain and the small domain is maintained in solution cannot be determined from these studies. The fact that there are two linkers between the domains may reduce the probability that the domains would separate in solution. As seen in Figure 2b, the linkers could conceivably act as a hinge to allow the XBM to swing out away from the CD. The substrate modeling studies presented here provide no insights into possible substrate−XBM binding because of the ambiguity of the interaction at the reducing end of the active site cleft. There are two aromatic residues (Tyr386 and Trp401) with significant solvent (and, hence, substrate) accessibility in the XBM. Either one or both of these residues may provide a binding platform for a carbohydrate substrate.

To our knowledge, the XBM module of XynA has not been classified. Table 10 lists the results of a BLAST search of the PIR-NREF database (*42*) using the last 94 amino acids of the XynA sequence. Nine unique proteins were found to match the XBM segment with levels of identity in the range of 31−84%, and all were identified as GH-5 enzymes with xylanase activity. This circumstantial evidence suggests that XBM is probably a xylan-binding module. Match 9 is only a fragment almost entirely composed of the XBM-like sequence. For the other eight matches, the results of the alignment of the catalytic domains with XynA are also included. For these eight matches, there is good alignment with both the XBM and the CD of XynA. Match 1 comes

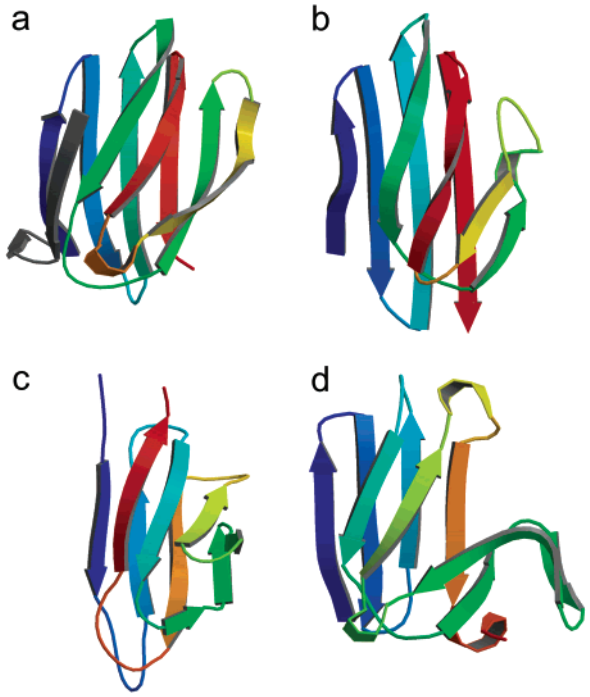


FIGURE 7: Representative carbohydrate-binding modules containing approximately 90 amino acids. In panel a is the XBM of XynA. In panel b is a representative of CBM-20 from *B. circulans* strain 251 (PDB entry 1CDG). In panel c is a representative of CBM-2 from *C. fimi* (PDB entry 2XBD), determined by NMR. In panel d is the CBM from pig pancreatic α-amylase (*78*). The structures in panels a, b, and d have a similar folding pattern, whereas that in panel c has a different pattern despite having approximately the same number of amino acids. This figure was rendered with the programs MOLSCRIPT (*63*) and RASTER3D (*64*).

from the same species as XynA, but from a different strain, which is clearly evident in the high level of identity between the two. Matches 2 and 3 obviously are similar to XynA, having approximately the same overall length; matches 4−8, on the other hand, appear to have at least a second C-terminal module of approximately 100−170 amino acids. In comparison to matches 4−7, match 8 appears to have a third C-terminal module. To our knowledge, none of the modules of matches 1−8 that align with the XBM have been classified in the CBM classification scheme.

We were able to identify three CBM families containing CBMs of approximately 100 residues, namely, CBM-2, CBM-20, and CBM-21. We found no three-dimensional structures for members of CBM-21. CBM-2 is represented only by NMR structures. Figure 7 offers a visual comparison between the XBM and representatives of CBM-2 (*76*) and CBM-20 (*77*), as well as porcine pancreatic α-amylase [PPA (*78*)] (whose CBM appears to remain unclassified at this time, but could very well belong to CBM-20). Coordinates of both the CBM-20 and CBM-2 representatives are from the Protein Data Bank (*46*). Clearly, XBM has the same topology as CBM-20 and PPA, while the CBM-2 model has a different topology.

*Conclusion.* The high-resolution three-dimensional structure of xylanase XynA from *E. chrysanthemi* has been determined. By analogy to GH-5 and GH-10 structures in the Protein Data Bank (*46*), the catalytic residues are identified as Glu165 and Glu253. Seven aromatic amino acids are involved in the formation of the active site binding cleft. The structural alignment data show that GH-10 xylanases make up a very closely related family with a high percentage of spatially equivalent and identical residues and small rms deviations between equivalent residues. The GH-5 glycoside hydrolases are a more diverse group of enzymes but still possess greater self-equivalency than is seen between GH-5 and GH-10 structures. XynA, although it has been classified as a family 5 glycoside hydrolase, is no more structurally equivalent to GH-5 than it is to GH-10, and in fact, it is approximately as equivalent to the GH-5 structures as the GH-5 structures are on average to the GH-10 structures.

## ACKNOWLEDGMENT

We acknowledge Mr. Aaron Greenwood for assistance in creating the illustrations.

## REFERENCES

1. Coutinho, P. M., and Henrissat, B. (1999) Carbohydrate-Active Enzymes server at http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html.
2. Coutinho, P. M., and Henrissat, B. (1999) in *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G. J., Henrissat, B., and Svensson, B., Eds.) pp 3−12, Royal Society of Chemistry, Cambridge, U.K.
3. Coutinho, P. M., and Henrissat, B. (1999) in *Genetics, Biochemistry and Ecology of Cellulose Degradation* (Ohmiya, K., Hayashi, K., Sakka, K., Kobayashi, Y., Karita, S., and Kimura, T., Eds.) pp 15−23, Uni Publishers Co., Tokyo.
4. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities, *Biochem. J. 280*, 309−316.
5. Henrissat, B., and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities, *Biochem. J. 293*, 781−788.
6. Henrissat, B., and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases, *Biochem. J. 316*, 695−696.
7. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases, *Structure 3*, 853−859.
8. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P., and Davies, G. (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases, *Proc. Natl. Acad. Sci. U.S.A. 92*, 7090−7094.
9. Henrissat, B., and Davies, G. (1997) Structural and sequence based classification of glycoside hydrolases, *Curr. Opin. Struct. Biol. 7*, 637−644.
10. Prade, R. A. (1993) in *Xylanases: from Biology to Biotechnology. Biotechnology and Genetic Engineering Reviews* (Tombs, M. P., Ed.) Vol. 13, pp 101−132, Intercept, Andover, MA.
11. Ryttersgaard, C., Lo Leggio, L., Coutinho, P. M., Henrissat, B., and Larsen, S. (2002) *Aspergillus aculeatus* β-1,4-galatanase: substrate recognition and relations to other glycoside hydrolases in clan GH-A, *Biochemistry 41*, 15135−15143.
12. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and Yaguchi, M. (1994) Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase, *Protein Sci. 3*, 467−475.
13. Wakarchuk, W. W., Sung, W. L., Campbell, R. L., Cunningham, A., Watson, D. C., and Yaguchi, M. (1994) Thermostabilization of the *Bacillus circulans* xylanase by the introduction of disulfide bond, *Protein Eng. 7*, 1379−1386.
14. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase, *Biochemistry 38*, 5346−5354.
15. Joshi, M. D., Sidhu, G., Pot, I., Brayer, G. D., Withers, S. G., and McIntosh, L. P. (2000) Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase, *J. Mol. Biol. 299*, 255−279.
16. Harris, G. W., Pickersgill, R. W., Connerton, I., Debeire, P., Touzel, J.-P., Breton, C., and Pérez, S. (1997) Structural basis of the properties of an industrially relevant thermophilic xylanase, *Proteins 29*, 77−86.
17. Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jorgensen, P. L., Schulein, M., Dupont, C., Davies, G. J., and Wilson, K. S.

(1999) Catalysis and specificity in enzymatic glycoside hydrolysis: A 2,5B conformation for the glycosyl-enzyme intermediate revealed by the structure of the *Bacillus agaradhaerens* family 11 xylanase, *Chem. Biol. 7*, 483−492.
18. Fushinobu, S., Ito, K., Konno, M., Wakagi, T., and Matsuzawa, H. (1998) Crystallographic and mutational analyses of an extremely acidophilic and acid-stable xylanase: biased distribution of acidic residues and importance of Asp37 for catalysis at low pH, *Protein Eng. 11*, 1121−1128.
19. Kumar, P. R., Eswaramoorthy, S., Vithayathil, P. J., and Viswamitra, M. A. (2000) The tertiary structure at 1.59 Å resolution and the proposed amino acid sequence of a family-11 xylanase from the thermophilic fungus *Paecilomyces varioti* Bainier, *J. Mol. Biol. 295*, 581−593.
20. Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W., and Kratky, C. (1998) Thermophilic xylanase from *Thermomyces lanuginosus*: high-resolution X-ray structure and modeling studies, *Biochemistry 37*, 13475−13485.
21. Törrönen, A., Harkki, A., and Rouvinen, J. (1994) Three-dimensional structure of endo-1,4-$\beta$-xylanase II from *Trichoderma reesei*: two conformational states in the active site, *EMBO J. 13*, 2493−2501.
22. Törrönen, A., and Rouvinen, J. (1995) Structural comparison of two major endo-1,4-xylanases from *Trichoderma reesei*, *Biochemistry 34*, 847−856.
23. Havukainen, R., Törrönen, A., Laitinen, T., and Rouvinen, J. (1996) Covalent binding of three epoxyalkyl xylosides to the active site of endo-1,4-xylanase II from *Trichoderma reesei*, *Biochemistry 35*, 9617−9624.
24. Krengel, U., and Dijkstra, B. W. (1996) Three-dimensional structure of endo-1,4-$\beta$-xylanase from *Aspergillus niger*: molecular basis for its low pH optimum, *J. Mol. Biol. 263*, 70−78.
25. White, A., Withers, S. G., Gilkes, N. R., and Rose, D. R. (1994) Crystal structure of the catalytic domain of the $\beta$-1,4-glycanase Cex from *Cellulomonas fimi*, *Biochemistry 33*, 12546−12552.
26. White, A., Tull, D., Johns, K., Withers, S. G., and Rose, D. R. (1996) Crystallographic observation of a covalent catalytic intermediate in a $\beta$-glycoside, *Nat. Struct. Biol. 3*, 149−154.
27. Notenboom, V., Birsan, C., Nitz, M., Rose, D. R., Warren, R. A. J., and Withers, S. G. (1998) Insight into transition state stabilization of the $\beta$-1,4-glycanase Cex by covalent intermediate accumulation in active site mutants, *Nat. Struct. Biol. 5*, 812−818.
28. Notenboom, V., Birsan, C., Warren, R. A. J., Withers, S. G., and Rose, D. R. (1998) Exploring the cellulose/xylan specificity of the $\beta$-1,4-glycanase Cex from *Cellulomonas fimi* through crystallography and mutation, *Biochemistry 37*, 4751−4758.
29. Notenboom, V., Williams, S. J., Hoos, R., Withers, S. G., and Rose, D. R. (2000) Detailed structural analysis of glycoside/inhibitor interactions: Complexes of Cex from *Cellulomonas fimi* with xylobiose-derived aza-sugars, *Biochemistry 39*, 11553−11563.
30. Harris, G. W., Jenkins, J. A., Connerton, I., Cummings, N., Lo Leggio, L., Scott, M., Hazlewood, G. P., Laurie, J. I., Gilbert, H. J., and Pickersgill, R. W. (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.
31. Harris, G. W., Jenkins, J. A., Connerton, I., and Pickersgill, R. W. (1996) Refined crystal structure of the catalytic core of xylanase A from *Pseudomonas fluorescens* at 1.8 Å resolution, *Acta Crystallogr. D52*, 393−401.
32. Derewenda, U., Swenson, L., Green, R., Wei, Y., Morosoli, R., Shareck, F., Kluepfel, D., and Derewenda, Z. (1994) Crystal structure, at 2.6-Å 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.
33. Ducros, V., Charnock, S. J., Derewenda, U., Derewenda, Z. S., Dauter, Z., Dupont, C., Shareck, F., Morosoli, R., Kluepfel, D., and Davies, G. J. (2000) Substrate specificity in glycoside hydrolase family 10. Structural and kinetic analysis of the *Streptomyces lividans* xylanase 10A, *J. Biol. Chem. 275*, 23020−23026.
34. Kaneko, S., Kuno, A., Fujimoto, Z., Shimizu, D., Machida, S., Sato, Y., Yura, K., Go, M., Mizuno, H., Taira, K., Kusakabe, I., and Hayashi, K. (1999) An investigation of the nature and function of module 10 in a family F/10 xylanase FXYN of *Streptomyces olivaceoviridis* E-86 by module shuffling with the Cex of *Cellulomonas fimi* and by site-directed mutagenesis, *FEBS Lett. 460*, 61−66.
35. Fujimoto, Z., Kuno, A., Kaneko, S., Yoshida, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2000) Crystal structure of *Streptomyces olivaceoviridis* E-86 $\beta$-xylanase containing xylan-binding domain, *J. Mol. Biol. 300*, 575−585.
36. Schmidt, A., Schlacher, A., Steiner, W., Schwab, H., and Kratky, C. (1998) Structure of the xylanase from *Penicillum simplicissimum*, *Protein Sci. 7*, 2081−2088.
37. Schmidt, A., Gübitz, G. M., and Kratky, C. (1999) Xylan binding subsite mapping in the xylanase from *Penicillum simplicissimum* using xylooligosaccharides as cryo-protectant, *Biochemistry 38*, 2403−2412.
38. Lo Leggio, L., Kalogiannis, S., Bhat, M. K., and Pickersgill, R. W. (1999) High-resolution structure and sequence of *T. aurantiacus* xylanase I: implications for the evolution of thermostability in family 10 xylanases and enzymes with $\beta/\alpha$-barrel architecture, *Proteins 36*, 295−306.
39. Natesh, R., Bhanumoorthy, P., Vithayathil, P. J., Sekar, K., Ramakumar, S., and Viswamitra, M. A. (1999) Crystal structure at 1.8 Å resolution and proposed amino acid sequence of a thermostable xylanase from *Thermoascus aurantiacus*, *J. Mol. Biol. 288*, 999−1012.
40. Teixeira, S., Lo Leggio, L., Pickersgill, R., and Cardin, C. (2001) Anisotropic Refinement of the Structure of *Thermoascus aurantiacus* Xylanase I, *Acta Crystallogr. D57*, 385−392.
41. Keen, N. T., Boyd, C., and Henrissat, B. (1996) Cloning and characterization of a xylanase gene from corn strains of *Erwinia chrysanthemi*, *Mol. Plant-Microbe Interact. 9*, 651−657.
42. Wu, C. H., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., Hu, Z.-Z., Ledley, R. S., Lewis, K. C., Mewes, H.-W., Orcutt, B. C., Suzek, B. E., Tsugita, A., Vinayaka, C. R., Yeh, L.-S. L., Zhang, J., and Barker, W. C. (2002) The Protein Information Resource: an integrated public resource of functional annotation of proteins, *Nucleic Acids Res. 30*, 35−37.
43. Braun, E. J., and Rodrigues, C. A. (1993) Purification and properties of an endoxylanase from a corn stalk rot strain of *E. chrysanthemi*, *Phytopathology 83*, 332−338.
44. Hurlbert, J. C., and Preston, J. F., III (2001) Functional characterization of a novel xylanase from a corn strain of *Erwinia chrysanthemi*, *J. Bacteriol. 183*, 2093−2100.
45. Barba de la Rosa, A. P., Day, J., Larson, S. B., Keen, N. T., and McPherson, A. (1997) Crystallization of xylanase from *Erwinia chrysanthemi*: influence of heat and polymeric substrate, *Acta Crystallogr. D53*, 256−261.
46. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The protein data bank, *Nucleic Acids Res. 28*, 235−242.
47. Terwilliger, T. C., and Eisenberg, D. (1983) Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions, *Acta Crystallogr. A39*, 813−817.
48. Terwilliger, T. C., Kim, S.-H., and Eisenberg, D. (1987) Generalized method of determining heavy-atom positions using the difference Patterson function, *Acta Crystallogr. A43*, 1−5.
49. Terwilliger, T. C., and Eisenberg, D. (1987) Isomorphous replacement: effects of errors on the phase probability distribution, *Acta Crystallogr. A43*, 6−13.
50. Terwilliger, T. C., and Berendzen, J. (1996) Correlated Phasing of Multiple Isomorphous Replacement Data, *Acta Crystallogr. D52*, 749−757.
51. Terwilliger, T. C. (1999) Automated MAD and MIR structure solution, *Acta Crystallogr. D55*, 849−861.
52. Furey, W. (1990) *PHASES*, American Crystallographic Association, Series 2 (18), Vol. 73.
53. Jones, T. A., Zou, J. Y., Cowan, S. W., and 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.
54. Kleywegt, G. J., and Jones, T. A. (1996) xdlMAPMAN and xdlDATAMAN: Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets, *Acta Crystallogr. D52*, 826−828.
55. Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) Crystallographic R factor refinement by molecular dynamics, *Science 235*, 458−460.
56. Brünger, A. T. (1991) Simulated annealing in crystallography, *Annu. Rev. Phys. Chem. 42*, 197−223.

57. Brünger, A. T. (1992) The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 427−474.
58. Sheldrick, G. M., and Schneider, T. R. (1997) SHELXL: High-resolution refinement, *Methods Enzymol. 277*, 319−343.
59. Luzzati, V. (1952) Traitement statistique des erreurs dans la determination des structure cristallines, *Acta Crystallogr. 5*, 802−810.
60. Evans, S. V. (1993) SETOR: hardware lighted three-dimensional solid model representations of macromolecules, *J. Mol. Graphics 11*, 134−138.
61. Ramachandran, G. N., and Sasisekharan, V. (1968) Conformation of polypeptides and proteins, *Adv. Protein Chem. 23*, 283−438.
62. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
63. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−950.
64. Merrit, E. A., and Murphy, M. E. P. (1994) Raster3D Version 2.0. A program for photorealistic molecular graphics, *Acta Crystallogr. D50*, 869−873.
65. Adobe Illustrator 8.0 (2003) Adobe Systems Inc.
66. Nicholls, A., Sharp, K., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins 11*, 281−296.
67. Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility, *J. Mol. Biol. 55*, 379−400.
68. Lo Leggio, L., and Larsen, S. (2002) The 1.62 Å structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5, *FEBS Lett. 523*, 103−108.
69. Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z., Wilson, K. S., Chauvaux, S., Beguin, P., and Alzari, P. M. (1995) A common protein fold and similar active site in two distinct families of $\beta$-glycanases, *Nat. Struct. Biol. 2*, 569−576.
70. Dominguez, R., Souchon, H., Lascombe, M., and Alzari, P. M. (1996) The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism, *J. Mol. Biol. 257*, 1042−1051.
71. Sabini, E., Schubert, H., Murshudov, G., Wilson, K. S., Siika-Aho, M., and Penttilä, M. (2000) The three-dimensional structure of a *Trichoderma reesei* $\beta$-mannanase from glycoside hydrolase family 5, *Acta Crystallogr. D56*, 3−13.
72. Varrot, A., Schulein, M., and Davies, G. J. (2000) Insights into ligand-induced conformational change in Cel5A from *Bacillus agaradhaerens* revealed by a catalytically active crystal form, *J. Mol. Biol. 297*, 819−828.
73. Davies, G. J., Dauter, M., Brzozowski, A. M., Bjornvad, M. E., Andersen, K. V., and Schulein, M. (1998) Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6 Å and its cellobiose complex at 2.0 Å resolution, *Biochemistry 37*, 1926−1932.
74. Wang, Q., Tull, D., Meinke, A., Gilkes, N. R., Warren, R. A. J., Aebersold, R., and Withers, S. G. (1993) Glu280 is the nucleophile in the active site of *Clostridium thermocellum* CelC, a family A endo-$\beta$-1,4-glucanase, *J. Biol. Chem. 268*, 14096−14102.
75. Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R., and Karplus, P. A. (1996) Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose, *Biochemistry 35*, 10648−10660.
76. Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J., and Williamson, M. P. (1999) A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity, *Structure 7*, 853−864.
77. Lawson, C. L., Van Montfort, R., Strokopytov, B., Rozeboom, H. J., Kalk, K. H., DeVries, G. E., Penninga, D., Dijkhuizen, L., and Dijkstra, B. W. (1994) Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form, *J. Mol. Biol. 236*, 590−600.
78. Larson, S. B., Greenwood, A., Cascio, D., Day, J., and McPherson, A. (1994) Refined molecular structure of pig pancreatic α-amylase at 2.1 Å resolution, *J. Mol. Biol. 235*, 1560−1584.

BI034144C