Acta Crystallographica Section F
**Structural Biology and Crystallization Communications**
ISSN 1744-3091

# Three-dimensional structure of a thermophilic family GH11 xylanase from *Thermobifida fusca*

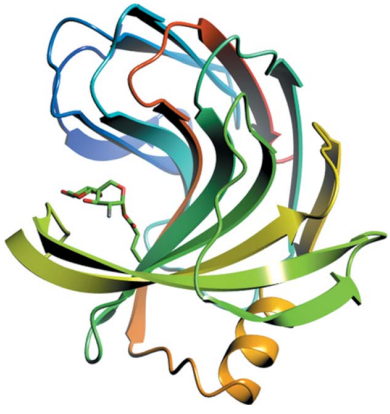
**Alicia Lammerts van Bueren,[a] Suzie Otani,[b] Esben P. Friis,[b] Keith S. Wilson[a] and Gideon J. Davies[a]***

[a]Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, England, and [b]Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark

Correspondence e-mail: davies@ysbl.york.ac.uk

Received 27 June 2011
Accepted 21 November 2011

**PDB Reference:** thermophilic family GH11 xylanase, 3zse.





Thermostable enzymes employ various structural features dictated at the amino-acid sequence level that allow them to maintain their integrity at higher temperatures. Many hypotheses as to the nature of thermal stability have been proposed, including optimized core hydrophobicity and an increase in charged surface residues to enhance polar solvent interactions for solubility. Here, the three-dimensional structure of the family GH11 xylanase from the moderate thermophile *Thermobifida fusca* in its trapped covalent glycosyl-enzyme intermediate complex is presented. Interactions with the bound ligand show fewer direct hydrogen bonds from ligand to protein than observed in previous complexes from other species and imply that binding of the xylan substrate involves several water-mediated hydrogen bonds.

## 1. Introduction

Xylanases are glycoside hydrolases that catalyse the hydrolysis of xylans, which are polymers of $\beta$-1,4-linked xylose residues and the major component of plant cell-wall hemicellulose. Glycoside hydrolases themselves (GH) have been classified into >118 families (http://www.cazy.org) to date on the basis of amino-acid sequence (Henrissat & Davies, 1997; Cantarel *et al.*, 2009); ten of the GH families are reported to contain xylanases (3, 8, 10, 11, 26, 39, 43, 52, 54 and 116), with the majority of endo-acting xylanases belonging to families GH 8, 10 and 11.

Xylanases are also model systems for the understanding of enzyme reaction mechanisms. GH11 xylanases follow a reaction mechanism with net retention of anomeric configuration (Gebler *et al.*, 1992; GH mechanisms have recently been reviewed in Vocadlo & Davies, 2008). In brief, they follow a double-displacement reaction (Koshland, 1953), in contrast to inverting enzymes, which hydrolyse glycosidic bonds in a single step. To date, the structures of over 20 different GH11 xylanases have been reported (see http://www.cazy.org/GH11_structure.html), all of which have a closely similar $\beta$-jelly-roll topology with conserved catalytic nucleophile and acid/base glutamate residues. Capturing snapshots along the reaction pathway of retaining GHs using X-ray crystallography has been extensively used to characterize their catalytic mechanism (for reviews, see Davies *et al.*, 2003, 2011). Most informative are the enzyme–substrate (or Michaelis) complexes and covalent glycosyl-enzyme intermediates which map out the electrophilic migration across the anomeric centre during catalysis. No Michaelis complexes are currently available for the GH11 family, but crystal structures of xylosyl-enzyme complexes suggest the formation of a covalent glycosyl intermediate in an unusual $^{2,5}B$ conformation during the glycosylation step (Sidhu *et al.*, 1999; Sabini *et al.*, 1999). Additional support for this pathway came from the observation of a similar $^{2,5}B$ conformation for a pseudo-product complex with a nucleophile variant (Sabini *et al.*, 2001). Indeed, pathways *via* unusual $^{2,5}B$ conformations may also pertain for GH6 and GH8 inverting cellulases (Davies *et al.*, 2011).

Xylanases have a wide range of industrial applications, such as bioenergy and biofuel production from plant cell-wall biomass and enhancement of animal feedstocks. Enzymes isolated from thermophilic bacteria are often favoured because they are better able to withstand the high temperatures that are incurred during the pro-

**Table 1**
Data-collection and refinement statistics for *T. fusca* GH11 (PDB entry 3zse).

| | |
|---|---|
| Data collection | |
| Space group | $P2_12_12_1$ |
| Unit-cell parameters (Å) | $a$ = 40.6, $b$ = 41.3, $c$ = 94.0 |
| Resolution | 45–1.78 (1.88–1.78) |
| $R_{merge}$ | 0.15 (0.50) |
| $\langle I/\sigma(I)\rangle$ | 6.4 (2.2) |
| Completeness (%) | 99.9 (99.9) |
| Multiplicity | 5.0 (5.0) |
| Refinement | |
| Resolution (Å) | 45–1.78 |
| No. of reflections | 78395 (11227) |
| Unique reflections | 15691 (2251) |
| $R_{work}/R_{free}$ | 0.19/0.24 |
| No. of atoms | |
| Protein | 1475 |
| Water | 146 |
| Ligands | |
| FXP | 18 |
| EDO | 24 |
| $B$ factors (Å$^2$) | |
| Protein | 16 |
| Water | 25 |
| Ligands | |
| FXP | 16 |
| EDO | 29 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.013 |
| Bond angles (°) | 1.39 |

duction process. Thermostability is achieved by structural changes, some of which are reflected in primary sequence, but whilst retaining the global three-dimensional fold. There is a considerable body of work across many enzyme systems attempting to both understand the relationship between sequence, structure and stability and to use such knowledge to engineer improved enzymes for the industrial milieu. Some of the many observations that may be linked to improved stability include a higher number of aromatic residues for increased compactness and a higher number of charged residues on the protein surface to increase water-mediated hydrogen-bonding networks (Vogt *et al.*, 1997; Zhou *et al.*, 2008; Kannan & Vishveshwara, 2000; Taylor & Vaisman, 2010).

Here, we report the structure of a thermostable xylanase (henceforth referred to as TfGH11) from *Thermobifida fusca*, a thermophilic soil bacterium which is a major contributor to the degradation of organic plant material in heated environments such as compost. The enzyme has been trapped in its covalent glycosyl-enzyme intermediate complex with the synthetic substrate 2,4-dinitrophenyl (DNP) 2-deoxy-2-fluoro-$\beta$-xylobioside (Ziser *et al.*, 1995). We have identified changes in the active-site residues in which several direct hydrogen bonds between enzyme and substrate are replaced by water-mediated hydrogen bonds. These water networks within the active site differ from previous observations of GH11 ligand complexes.

## 2. Crystallization and structure determination

Purified TfGH11 was provided as a gift from Novozymes A/S (Bagsvaerd, Denmark) in 25 m*M* acetic acid pH 5.0, 100 m*M* NaCl. Prior to crystallization, 200 m*M* potassium citrate and solid DNP 2-deoxy-2-fluoro-$\beta$-xylobioside were added to an 8 mg ml$^{-1}$ enzyme solution. Initially, a brown precipitate formed in 1 d, from which nucleation and growth of crystals occurred. The best crystals (diamond-shaped; ~0.2 × 0.2 × 0.25 µm) formed in about 4 d from a precipitant solution consisting of 0.2 *M* $CaCl_2$, 15% PEG 3350, 8%

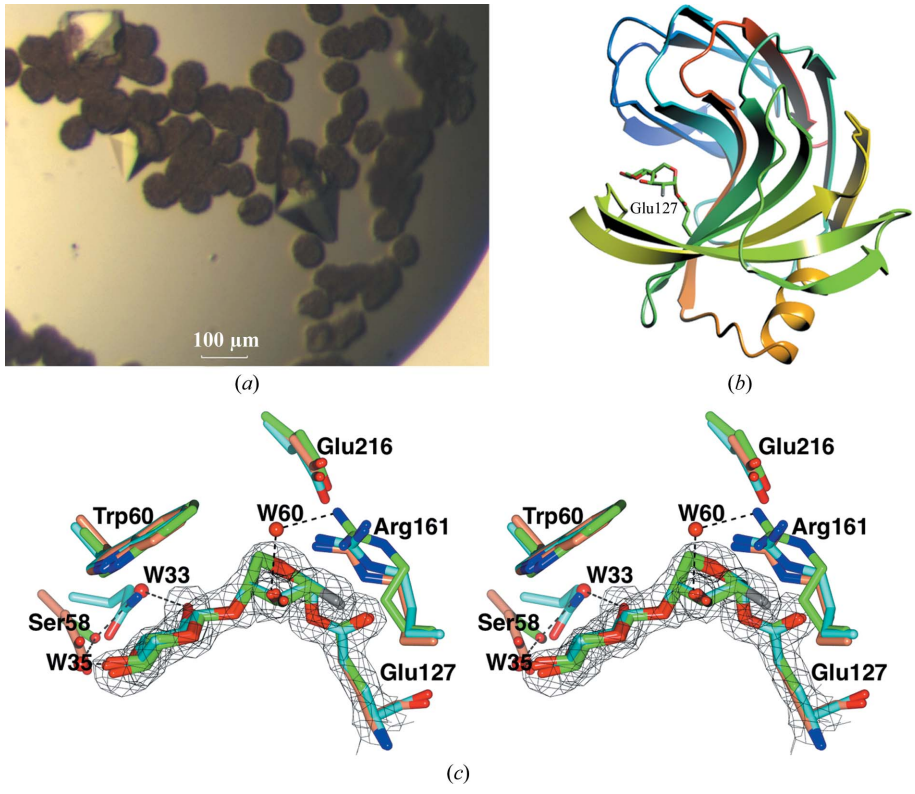



**Figure 1**
(*a*) Crystals of the TfGH11–DNP 2-deoxy-2-fluoro-$\beta$-xylobiose complex that arise out of the precipitate. (*b*) $\beta$-Jelly-roll topology of TfGH11 shown as ribbons. The catalytic nucleophile (Glu127) and covalent 2-fluoro-xylobiosyl-enzyme intermediate are shown as sticks in green. (*c*) Observed electron density (*REFMAC* weighted $2F_{obs} - F_{calc}$ at 1$\sigma$) for the 2-fluoro-xylobiosyl-enzyme intermediate (green) in divergent stereo. Surrounding residues are shown, along with the overlaid *B. agaradhaerens* GH11 (blue; PDB entry 1qh6) and *B. circulans* GH11 (yellow; PDB entry 1bvv) intermediate complexes. (*b*) and (*c*) were drawn with *CCP4mg* (McNicholas *et al.*, 2011).

PEG 550 MME and 0.1 *M* sodium acetate pH 4.6 (Fig. 1*a*). Crystals were cryoprotected in mother liquor supplemented with 20% ethylene glycol and flash-cooled in liquid $N_2$.

Data from a single crystal were collected on beamline I04 at the Diamond Light Source (Oxford, England) to 1.78 Å resolution and were initially, and erroneously, indexed as tetragonal *P*4/*mmm*, with unit-cell parameters $a = b = 41.1$, $c = 93.5$ Å. The data were integrated in *iMOSFLM* (Leslie, 1992) with subsequent computations using the *CCP*4 suite (Winn *et al.*, 2011). The correct space group determined with *POINTLESS* was actually $P2_12_12_1$, with final unit-cell parameters $a = 40.6$, $b = 41.3$, $c = 94.0$ Å, and the data were merged with *SCALA* in this space group prior to structure solution and refinement.

The structure was solved by molecular replacement with the *CCP*4 version of *MOLREP* (Vagin & Teplyakov, 2010) using the xylanase from *Thermopolyspora flexulosa* (PDB entry 1m4w) with 85.5% sequence identity (Hakulinen *et al.*, 2003) as a search model. A clear solution with one molecule per asymmetric unit led to $R$ and $R_{\rm free}$ values of 0.28 and 0.33, respectively, after a single round of refinement with *REFMAC* (Murshudov *et al.*, 2011). A complete model was built with *Coot* (Emsley & Cowtan, 2004) which contained amino acids 43–230 of TfGH11. The starting coordinates for the 2-fluoro-xylobioside substrate intermediate were taken from the *Bacillus agaradhaerens* GH11 PDB file 1qh6 (Sabini *et al.*, 1999) and a refinement library was generated using the *PRODRG* online server (Schüttelkopf & van Aalten, 2004; http://davapc1.bioch.dundee.ac.uk/prodrg/). Data-collection and refinement statistics are given in Table 1.

## 3. Structural features of *T. fusca* xylanase

The final TfGH11 model has crystallographic $R$ and $R_{\rm free}$ values of 0.19 and 0.24, respectively, and contains amino acids 43–230 with a solvent content of 35%, with 1475 protein atoms (average $B$ value 15 Å$^2$), 18 ligand (covalent intermediate) atoms (average $B$ value 16 Å$^2$) and 146 water molecules (average $B$ value 25 Å$^2$). The structure is very similar to that of the search model (PDB entry 1m4w), with 85% sequence identity and an r.m.s.d. of 0.55 Å in $C^\alpha$ positions for all 187 modelled residues after structure superposition.

TfGH11 is a 20 kDa protein which adopts the classic $\beta$-jelly-roll fold seen in all other members of this family (for examples, see Wakarchuk *et al.*, 1994; Sabini *et al.*, 1999; Fig. 1*b*). It contains two $\beta$-sheets ($\beta$A and $\beta$B) with five and eight strands, respectively. $\beta$B forms a groove which houses the active site. We believe that TfGH11 is the first thermostable xylanase to be solved in complex with its trapped covalent-intermediate state. The catalytic nucleophile Glu127 is covalently linked to the 2-deoxy-2-fluoro-xylosyl moiety occupying the −1 subsite, while the sugar ring unambiguously adopts a $^{2,5}B$ distorted ring conformation, as previously shown for other GH11 enzymes (Sidhu *et al.*, 1999; Sabini *et al.*, 1999). However, the interpretation of intermediate complexes for GH11 is speculative because a Michaelis complex has yet to be reported for this family, leaving open the question as to whether these intermediate complexes represent a mechanistically relevant species. In the absence of a catalytically relevant Michaelis complex, a recent modelling study on the *B. circulans* GH11 xylanase concluded that the Michaelis complex is (also) in a $^{2,5}B$ conformation (Soliman *et al.*, 2009), suggesting that the distorted forms observed in experimental three-dimensional structures are relevant to the genuine catalytic pathway.

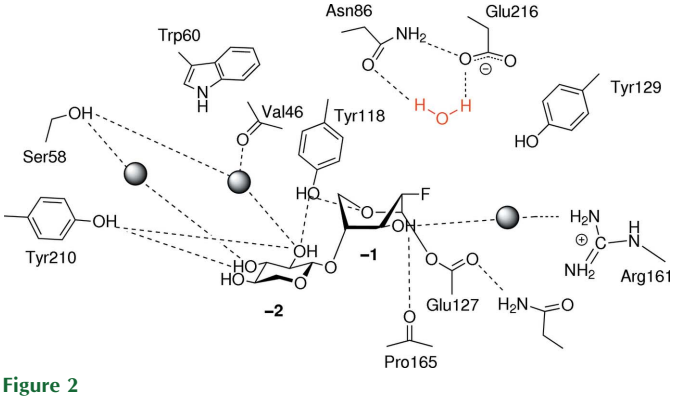



**Figure 2**
Schematic diagram of the interactions of the 2-fluoro-xylobiosyl-enzyme intermediate of TfXyn11. Water molecules are shown as shaded spheres, with the exception of the putative nucleophilic water, which is shown in red. Possible hydrogen bonds of <3.0 Å are shown as dotted lines.

Directly adjacent to the nucleophile Glu127 is the acid/base residue Glu216, which is hydrogen bonded to a water molecule positioned for in-line attack at the anomeric carbon centre of 2-fluoro-xylose in the −1 subsite. Additionally, there are several other contacts between the two xylose rings and the enzyme that are conserved in all GH11 enzymes (Figs. 1*c* and 2). One aromatic residue, aromatic Trp60, forms hydrophobic stacking against xylose in the −2 subsite. As in most carbohydrate-binding proteins, these C—H $\pi$ interactions are crucial for substrate binding (see, for example, Rudsander *et al.*, 2008; Vyas, 1991). Additionally, there are four direct hydrogen bonds, two between −2 xylose and the Tyr210 3′OH and Tyr118 2′OH groups and two between −1 xylose 3′OH and the Tyr118 endocyclic oxygen and Pro165 carboxyl main-chain group, that stabilize the complex (Fig. 1*c*).

There are three additional water-mediated hydrogen bonds in TfGH11 that replace conserved direct hydrogen bonds in other GH11 enzymes reported to date (Figs. 1*c* and 2). The closely related enzymes from *B. circulans* (BcxGH11; PDB entry 1bvv; Sidhu *et al.*, 1999) and *B. agaradhaerens* (BadXGH11; PDB entry 1qh6; Sabini *et al.*, 1999) share 69 and 55% sequence identity with TfGH11, respectively (r.m.s.d. of 0.75 Å over 171 aligned residues and r.m.s.d. of 0.88 Å over 183 residues, respectively). Comparison of the known glycosyl-enzyme intermediate structures of these closely similar enzymes with TfGH11 allowed us to identify changes within the active site that may aid catalysis at higher temperatures.

## 4. Water-mediated hydrogen bonds in substrate binding

Many authors have tried to correlate features of primary sequence that differ in proteins from thermophilic organisms in an attempt to gain some insight into factors that may influence protein stability. For example, amino-acid sequence characteristics such as a higher proportion of charged residues and a higher number of aromatic residues are all thought to contribute increased compactness and stability in thermostable enzymes (Vogt *et al.*, 1997; Taylor & Vaisman, 2010). Indeed, TfGH11 has a high number of aromatic residues (30/187; 16%). Additionally, thermostable enzymes often contain an unusually large number of surface polar residues which generate more water-mediated networks surrounding the enzyme; however, it is not apparent whether these networks affect the active-site surface of the enzymes.

In our structure, changes within the active site of TfGH11 lead to significant differences compared with those of previously deposited Xyn11 structures *via* the introduction of three water-mediated hydrogen bonds when compared with the known xylanase structures BadXGH11 and BcxGH11 (Figs. 1*c* and 2). The conserved Arg161 is

rotated −90°, causing the side-chain guanidinium group to lie perpendicular to the active site. This movement opens up a space for a water molecule, replacing the direct hydrogen bond seen in other GH11s with a water-mediated hydrogen bond between Arg161 NH2 and the 3′OH of the −1 xylose. It is not clear why this rotation occurs, as no additional interactions are made with Arg161 in TfGH11. A second change alters the hydrogen-bonding pattern with xylose in the −2 subsite, in which Glu17 (BadXGH11) or Gln7 (BcxGH11) is replaced by Ser58 in TfGH11. Serine, which has a shorter side chain than Glu and Gln, allows two water molecules to fill the excess space, forming two water-mediated hydrogen bonds with the 2′OH and 3′OH groups of the xylose monomer.

Directed evolution of xylanases has shown that more thermostable xylanases can be created when the resulting amino-acid substitutions increase the internal hydrophobicity of the enzyme (Miyazaki *et al.*, 2006; Xie *et al.*, 2006) and change the number of surface protein–solvent interactions (Ruller *et al.*, 2008), and this rationale has also been applied to other systems including lipases (Ahmad *et al.*, 2008). For TfGH11 we observe incorporation of water-mediated hydrogen bonds between enzyme and substrate; one may speculate that such networks somehow contribute to efficient catalysis by the *T. fusca* enzyme in a manner different from that observed in past xylanase complexes. Whether such interactions contribute to maintenance of activity at higher temperatures warrants further investigation.

GJD is a Royal Society/Wolfson Research Merit Award recipient and receives funding for glycosidase research from the BBSRC. KSW thanks Novozymes A/S for support. ALvB is funded by NSERC and EMBO postdoctoral fellowships.

## References

Ahmad, S., Kamal, M. Z., Sankaranarayanan, R. & Rao, N. M. (2008). *J. Mol. Biol.* **381**, 324–340.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Davies, G. J., Ducros, V. M., Varrot, A. & Zechel, D. L. (2003). *Biochem. Soc. Trans.* **31**, 523–527.
Davies, G. J., Planas, A. & Rovira, C. (2011). *Acc. Chem. Res.*, doi:10.1021/ar2001765.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gebler, J., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Béguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Warren, R. A. & Withers, S. G. (1992). *J. Biol. Chem.* **267**, 12559–12561.
Hakulinen, N., Turunen, O., Janis, J., Leisola, M. & Rouvinen, J. (2003). *Eur. J. Biochem.* **270**, 1399–1412.
Henrissat, B. & Davies, G. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.
Kannan, N. & Vishveshwara, S. (2000). *Protein Eng.* **13**, 753–761.
Koshland, D. E. (1953). *Biol. Rev.* **28**, 416–436.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.
Miyazaki, K., Takenouchi, M., Kondo, H., Noro, N., Suzuki, M. & Tsuda, S. (2006). *J. Biol. Chem.* **281**, 10236–10242.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Rudsander, U. J., Sandstrom, C., Piens, K., Master, E. R., Wilson, D. B., Brumer, H. III, Kenne, L. & Teeri, T. T. (2008). *Biochemistry*, **47**, 5235–5241.
Ruller, R., Deliberto, L., Ferreira, T. L. & Ward, R. J. (2008). *Proteins*, **70**, 1280–1293.
Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jørgensen, P. L., Schülein, M., Dupont, C., Davies, G. J. & Wilson, K. S. (1999). *Chem. Biol.* **6**, 483–492.
Sabini, E., Wilson, K. S., Danielsen, S., Schülein, M. & Davies, G. J. (2001). *Acta Cryst.* D**57**, 1344–1347.
Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Cryst.* D**60**, 1355–1363.
Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L. & Brayer, G. D. (1999). *Biochemistry*, **38**, 5346–5354.
Soliman, M. E., Ruggiero, G. D., Pernía, J. J., Greig, I. R. & Williams, I. H. (2009). *Org. Biomol. Chem.* **7**, 460–468.
Taylor, T. J. & Vaisman, I. I. (2010). *BMC Struct. Biol.* **10**, S5.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Vocadlo, D. J. & Davies, G. J. (2008). *Curr. Opin. Chem. Biol.* **12**, 539–555.
Vogt, G., Woell, S. & Argos, P. (1997). *J. Mol. Biol.* **269**, 631–643.
Vyas, N. K. (1991). *Curr. Opin. Struct. Biol.* **1**, 732–740.
Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J. & Yaguchi, M. (1994). *Protein Sci.* **3**, 467–475.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xie, H., Flint, J., Vardakou, M., Lakey, J. H., Lewis, R. J., Gilbert, H. J. & Dumon, C. (2006). *J. Mol. Biol.* **360**, 157–167.
Zhou, X.-X., Wang, Y.-B., Pan, Y.-J. & Li, W.-F. (2008). *Amino Acids*, **34**, 25–33.
Ziser, L., Setyawati, I. & Withers, S. G. (1995). *Carbohydr. Res.* **274**, 137–153.