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

Structure of *Acidothermus cellulolyticus* family 74 glycoside hydrolase at 1.82 Å resolution

Markus Alahuhta, William S. Adney, Michael E. Himmel and Vladimir V. Lunin*

BioSciences Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401, USA

Correspondence e-mail: vladimir.lunin@nrel.gov

Received 27 September 2013
Accepted 1 November 2013

PDB reference: family 74 glycoside hydrolase, 4lgn

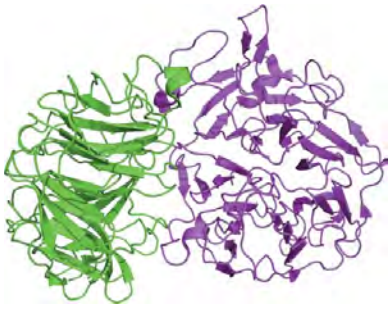
Here, a 1.82 Å resolution X-ray structure of a glycoside hydrolase family 74 (GH74) enzyme from *Acidothermus cellulolyticus* is reported. The resulting structure was refined to an R factor of 0.150 and an $R_{\rm free}$ of 0.196. Structural analysis shows that five related structures have been reported with a secondary-structure similarity of between 75 and 89%. The five similar structures were all either *Clostridium thermocellum* or *Geotrichum* sp. M128 GH74 xyloglucanases. Structural analysis indicates that the *A. cellulolyticus* GH74 enzyme is an endoxyloglucanase, as it lacks a characteristic loop that blocks one end of the active site in exoxyloglucanases. Superimposition with the *C. thermocellum* GH74 shows that Asp451 and Asp38 are the catalytic residues.

1. Introduction

Xyloglucan is an important polymer found in plant cell walls, where it facilitates the tightening and loosening of cellulose microfibrils, enabling cells to change shape during growth. It is also thought to rigidify the final shape of the cell wall by cross-linking adjacent cellulose microfibrils (Hayashi, 1989; Hayashi & Kaida, 2011). Xyloglucan primarily consists of repeating units of either XXXG or XXGG, where G is unbranched β-1→4-linked D-glucopyranose (glucan) and X refers to β-D-glucopyranose modified by an α-D-xylopyranose linked through a 1→6 bond (Vincken *et al.*, 1997). Depending on the plant species, the α-D-xylopyranose can be replaced or further linked/branched to other sugars such as galactose, arabinose or fucose (Fry, 1989).

Xyloglucan is degraded by xyloglucanases, most of which belong to glycosyl hydrolase family 74 (GH74) according to the CAZy classification (Cantarel *et al.*, 2009). Three GH74 xyloglucanases have been structurally characterized to date: a reducing-end-specific exo-GH74 and an endo-GH74 from *Geotrichum* sp. M128 (*G*GH74) (Yaoi *et al.*, 2004, 2007, 2009) and an endo-GH74 from *C. thermocellum* (*Ct*GH74; Martinez-Fleites *et al.*, 2006). The GH74 active site is typically an open groove, whereas the exo-GH74 displays a specific loop blocking one side of the cleft. In the endo-GH74s this loop is missing, allowing the enzyme to bind to the middle of the substrate (Yaoi *et al.*, 2007). GH74 enzymes catalyse xyloglucan depolymerization using a concerted hydrolysis mechanism which inverts the anomeric configuration of the product hydroxyl. In this mechanism the catalytic acid facilitates the departure of a leaving group by protonation. The catalytic base is thought to activate a water molecule by deprotonation (McCarter & Withers, 1994). Both the acid and the base have been proposed to be aspartic acid residues in GH74 enzymes (Yaoi *et al.*, 2004; Martinez-Fleites *et al.*, 2006).

Xyloglucan is a key hemicellulose in the growing cell wall of dicotyledonous plants. It links cellulose microfibrils together; therefore, its removal should improve the accessibility to cellulose. In this paper, we report the X-ray structure of an *A. cellulolyticus* thermally tolerant GH74 (*Ac*GH74; Ding *et al.*, 2003, 2008). The current work is the starting point for future studies of the importance of *Ac*GH74 in the degradation of cellulosic biomass and its structure–function relationship.



2. Materials and methods

2.1. Cloning, expression and purification

*Ac*GH74 was cloned from *A. cellulolyticus* as described in example 1 of Ding *et al.* (2008). The purified catalytic module was obtained using pET-22b (Novagen, Madison, Wisconsin, USA) containing the *Ac*GH74 sequence, which was introduced into *Escherichia coli* BL21 (DE3) competent cells by heat-shock transformation. Luria broth cultures with 0.1 µg ml^{-1} ampicillin were grown to mid-log phase at 310 K. The cultures were cooled slowly to 290 K, induced with 750 µ*M* IPTG and incubated overnight. The cells were harvested by centrifugation, frozen at 253 K and lysed with BugBuster HT (EMD Millipore, Merck KGaA, Darmstadt, Germany) according to the manufacturer's protocols. Cleared lysates were passed through a His-Select column (Sigma–Aldrich, St Louis, Missouri, USA) in 20 m*M* phosphate buffer, 300 m*M* NaCl, 10 m*M* imidazole pH 8.0 and eluted in the same buffer containing 250 m*M* imidazole. Eluted proteins were further purified by hydrophobic interaction chromatography using a Tricorn 10/100 column packed with Source 15PHE (GE Healthcare, Little Chalfont, England) using 20 m*M* bis-tris pH 6.5 with a descending 2 *M* ammonium sulfate gradient followed by gel filtration on a 26/60 Superdex 75 column using 20 m*M* acetate, 100 m*M* NaCl pH 5.0. The purity was assessed by SDS–PAGE and the concentration was determined using the BCA protein assay (Thermo Scientific Pierce Protein Biology Products, Rockford, Illinois, USA).

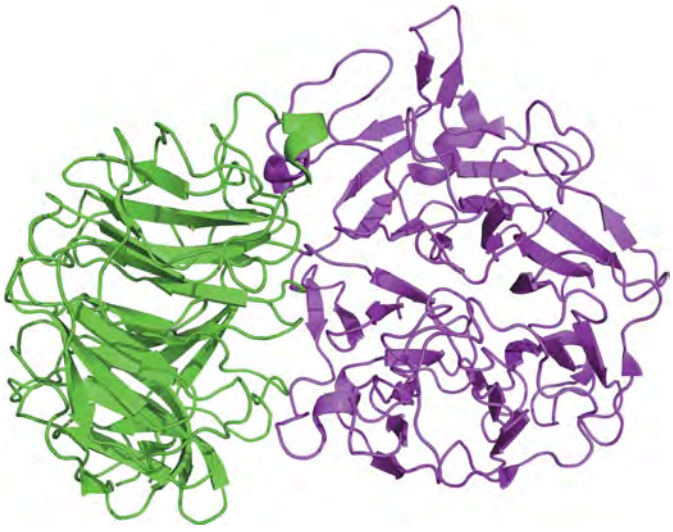


Figure 1
(*a*) The overall fold of *Ac*GH74 with the active-site cleft in the middle. The two interconnected domains are shown in green and magenta.

2.2. Crystallization

*Ac*GH74 crystals were obtained by hanging-drop vapor diffusion using a 24-well plate. 500 µl well solution was added to the reservoirs and the drops consisted of 1 µl well solution and 1 µl protein solution. The crystals were grown at 293 K with 0.1 m*M* trisodium dihydrate pH 5, 1.5 *M* sodium formate as the well solution. The protein solution consisted of 1.45 mg ml^{-1} protein in 20 m*M* acetate pH 5, 100 m*M* NaCl.

2.3. Data collection and processing

The *Ac*GH74 crystal was flash-cooled in a nitrogen-gas stream at 100 K before data collection using an in-house Bruker X8 MicroStar X-ray generator with Helios mirrors and a Bruker PLATINUM[135] CCD detector. Before cooling, the crystal was briefly submersed in the crystallization solution with 10%(*v*/*v*) ethylene glycol and 10%(*v*/*v*) glycerol as cryoprotectants. Data were indexed and

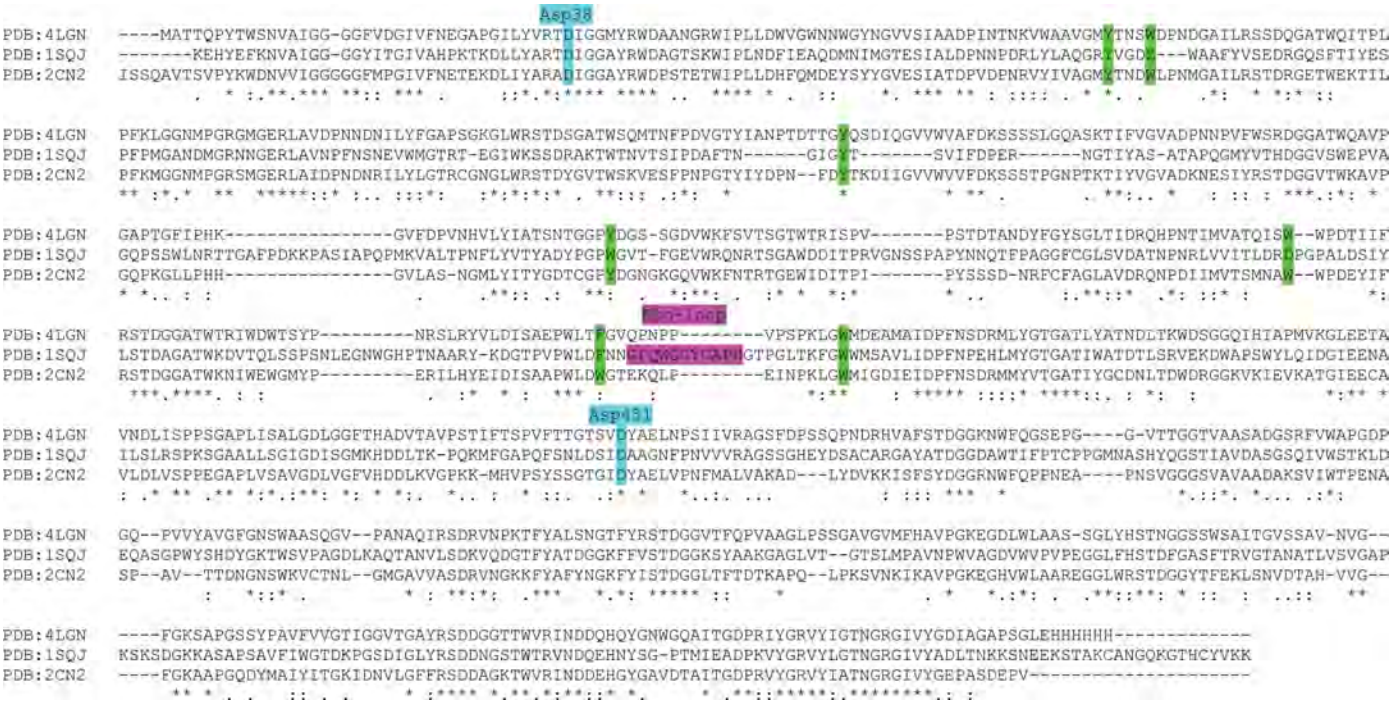


Figure 2
Sequence alignment of *Ac*GH74 (PDB entry 4lgn), *Ct*GH74 (PDB entry 2cn2; Martinez-Fleites *et al.*, 2006) and *G*GH74 (PDB entry 1sqj; Yaoi *et al.*, 2004). The catalytic residues are highlighted in cyan and the exo-loop of *G*GH74 in shown in red. The aromatic residues stacking with the substrate (according to PDB entry 2cn3; Martinez-Fleites *et al.*, 2006) are highlighted in green.

processed with the Bruker Suite of programs v.2008.1-0 (Bruker AXS, Madison, Wisconsin, USA).

2.4. Structure solution, refinement and structure analysis

Intensities were converted to structure factors and 5% of the reflections were flagged for R_{free} calculations using *F2MTZ*, *CTRUNCATE*, *CAD* and *UNIQUE* from the *CCP*4 suite (Winn *et al.*, 2011). *MOLREP* v.10.2.23 (Vagin & Teplyakov, 2010) was used for molecular replacement using the *C. thermocellum* family 74 xyloglucanase (PDB entry 2cn3; Martinez-Fleites *et al.*, 2006) as the search model. Refinement and manual corrections were performed using *REFMAC*5 v.5.7.0032 (Murshudov *et al.*, 2011) and *Coot* v.0.6.2 (Emsley *et al.*, 2010). *MolProbity* (Chen *et al.*, 2010) was used to analyze the Ramachandran plot, and root-mean-square deviations (r.m.s.d.s) of bond lengths and angles were calculated using the Engh and Huber ideal values of stereochemical parameters (Engh & Huber, 1991). Average *B* factors were calculated using *ICM* v.3.7-2e (Molsoft LLC, La Jolla, California, USA). *Coot* and *PyMOL* (http://www.pymol.org) were used to compare and analyse structures. Figs. 1 and 3 were created using *PyMOL*. Fig. 2 was created using *Clustal Omega* (Sievers *et al.*, 2011). The data-collection and refinement statistics are shown in Table 1.

3. Results and discussion

3.1. The structure of *Ac*GH74

We have solved a 1.82 Å resolution X-ray structure of an *A. cellulolyticus* GH74 enzyme. The resulting structure was refined to an *R* factor of 0.150 and an R_{free} of 0.196 with one molecule in the asymmetric unit. The overall fold of *Ac*GH74 contains tandem seven-bladed β-propeller N- and C-terminal domains (Fig. 1). The N-terminal domain is oriented at an angle of approximately 90° relative to the C-terminal domain. The structure has been deposited in the Protein Data Bank as entry 4lgn.

3.2. Comparison to other family 74 glycoside hydrolases

Pairwise secondary-structure matching by *PDBeFold* (Krissinel & Henrick, 2004) found five structures with a secondary-structure similarity of between 75 and 89%. The most similar structures were those of the *C. thermocellum* GH74 xyloglucanase (PDB entries 2cn3 and 2cn2; Martinez-Fleites *et al.*, 2006), with an r.m.s.d. of 1.08 Å and a sequence identity of 54% from *PDBeFold*. The other three similar structures were GH74 xyloglucanases from *Geotrichum* sp. M128 (PDB entries 3a0f, 1sqj and 2ebs; Yaoi *et al.*, 2004, 2007, 2009), with sequence identities of between 35 and 38% (Fig. 2).

A closer comparison with the other GH74 enzymes shows that the overall fold of *Ac*GH74 is highly similar to those of *G*GH74 and *Ct*GH74, except for a loop between residues 375 and 385 in the exo *G*GH74 (PDB entry 1sqj) that is missing from the others (Figs. 2 and 3). This well characterized loop has been shown to be the difference

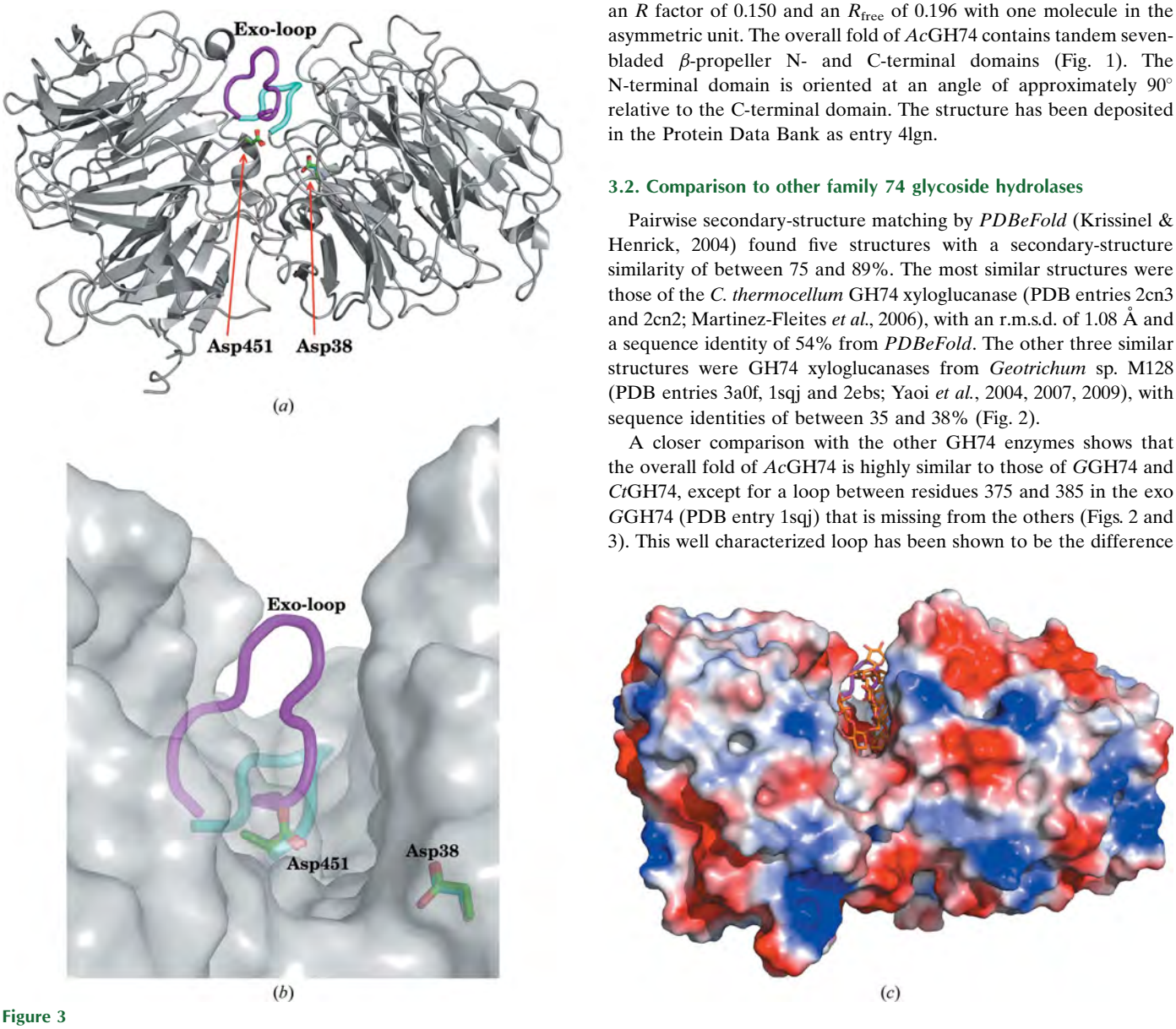


Figure 3
(*a*) A view along the active-site cleft of *Ac*GH74. (*b*) Enlarged view of the active site. (*c*) The electrostatic surface of *Ac*GH74 with a view along the active-site cleft showing the blocking exo-loop of *G*GH74 (PDB entry 1sqj) and the xyloglucan substrate of *Ct*GH74 (orange C atoms and red O atoms; PDB entry 2cn3). The exo-loop of *G*GH74, which does not exist in *Ac*GH74, is shown as a magenta ribbon. The corresponding loop in *Ac*GH74 is highlighted as a cyan ribbon in (*a*) and (*b*). The rest of *G*GH74 superimposed with *Ac*GH74 is not shown. *Ac*GH74 is shown as a grey ribbon. The side chains of the catalytic residues Asp480 (left) and Asp70 (right) of *Ct*GH74 are shown as sticks with green C atoms and red O atoms, and the side chains of the proposed catalytic residues Asp451 (left) and Asp38 (right) of *Ac*GH74 are shown as sticks with cyan C atoms and red O atoms.

Table 1
X-ray data-collection and refinement statistics.

Values in parentheses are for the highest resolution bin.

Data collection	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 74.8, b = 78.7, c = 144.0
Wavelength (Å)	1.54178
Temperature (K)	100
Resolution (Å)	25–1.82 (1.91–1.82)
Unique reflections	76875 (10202)
Observed reflections	435112 (44276)
R_{int}†	0.1557 (0.5343)
Average multiplicity	5.6 (4.3)
$\langle I\rangle/\langle\sigma(I)\rangle$	8.52 (2.0)
Completeness (%)	100 (100)
Refinement	
R/R_{free}	0.150 (0.265)/0.196 (0.320)
No. of protein atoms	5624
No. of water molecules	1159
No. of other atoms	125
R.m.s.d. from ideal bond lengths‡ (Å)	0.020
R.m.s.d. from ideal bond angles‡ (°)	1.9
Wilson B factor (Å^2)	11.5
Average B factor for protein atoms (Å^2)	10.1
Average B factor for water molecules (Å^2)	23.3
Ramachandran plot statistics§ (%)	
Allowed	100
Favored	97.2
Outliers	0

† $R_{int} = \sum|I - \langle I\rangle|/\sum|I|$, where I is the intensity of an individual reflection, $\langle I\rangle$ is the mean intensity of a group of equivalents and the sums are calculated over all reflections with more than one equivalent measured. ‡ Engh & Huber (1991). § Lovell *et al.* (2003).

between exo-GH74s and endo-GH74s (Yaoi *et al.*, 2007, 2009). Also, aspartate residues 38 and 451 in *Ac*GH74 perfectly superimpose with aspartate residues 70 and 480 in *Ct*GH74, which have been proposed to be the catalytic residues in GH74 enzymes (Figs. 2 and 3; Martinez-Fleites *et al.*, 2006). Since the exo-loop is clearly missing from *Ac*GH74, we structurally classify it as an endo-GH74 with Asp38 and Asp451 as the catalytic residues.

This work was funded in part by the US DOE Office of Science, Biological and Environmental Research Program, Bioenergy Research Center (BioEnergy Science Center, BESC) managed by Oak Ridge National Laboratory. The work was also funded by the US DOE BioEnergy Technology Office (BETO) in EERE.

References

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Ding, S.-Y., Adney, W. S., Vinzant, T. B., Decker, S. R., Baker, J. O., Thomas, S. R. & Himmel, M. E. (2003). *ACS Symp. Ser.* **855**, 332–360.
Ding, S.-Y., Adney, W. S., Vinzant, T. B. & Himmel, M. E. (2008). US Patent 7364890.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Fry, S. C. (1989). *J. Exp. Bot.* **40**, 1–11.
Hayashi, T. (1989). *Annu. Rev. Plant Physiol.* **40**, 139–168.
Hayashi, T. & Kaida, R. (2011). *Mol. Plant*, **4**, 17–24.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
Martinez-Fleites, C., Guerreiro, C. I. P. D., Baumann, M. J., Taylor, E. J., Prates, J. A. M., Ferreira, L. M. A., Fontes, C. M. G. A., Brumer, H. & Davies, G. J. (2006). *J. Biol. Chem.* **281**, 24922–24933.
McCarter, J. D. & Withers, S. G. (1994). *Curr. Opin. Struct. Biol.* **4**, 885–892.
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.
Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). *Mol. Syst. Biol.* **7**, 539.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Vincken, J.-P., York, W. S., Beldman, G. & Voragen, A. G. J. (1997). *Plant Physiol.* **114**, 9–13.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yaoi, K., Kondo, H., Hiyoshi, A., Noro, N., Sugimoto, H., Tsuda, S., Mitsuishi, Y. & Miyazaki, K. (2007). *J. Mol. Biol.* **370**, 53–62.
Yaoi, K., Kondo, H., Hiyoshi, A., Noro, N., Sugimoto, H., Tsuda, S. & Miyazaki, K. (2009). *FEBS J.* **276**, 5094–5100.
Yaoi, K., Kondo, H., Noro, N., Suzuki, M., Tsuda, S. & Mitsuishi, Y. (2004). *Structure*, **12**, 1209–1217.