Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

X-ray crystallographic native sulfur SAD structure determination of laminarinase Lam16A from *Phanerochaete chrysosporium*

Jonas Vasur,[a] Rie Kawai,[b] Anna M. Larsson,[c] Kiyohiko Igarashi,[b] Mats Sandgren,[a] Masahiro Samejima[b] and Jerry Ståhlberg[a]*

[a]Department of Molecular Biology, Swedish University of Agricultural Sciences, PO Box 590, SE-75124 Uppsala, Sweden, [b]Department of Biomaterial Sciences, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan, and [c]Department of Cell and Molecular Biology, Uppsala University, PO Box 596, SE-75124 Uppsala, Sweden

Correspondence e-mail: jerry.stahlberg@molbio.slu.se

Received 29 July 2006
Accepted 8 September 2006

PDB Reference: laminarinase Lam16A, 2cl2, r2cl2sf.

Laminarinase Lam16A from *Phanerochaete chrysosporium* was recombinantly expressed in *Pichia pastoris*, crystallized and the structure was solved at 1.34 Å resolution using native sulfur SAD X-ray crystallography. It is the first structure of a non-specific 1,3(4)-β-D-glucanase from glycoside hydrolase family 16 (GH16). *P. chrysosporium* is a wood-degrading basidiomycete fungus and Lam16A is the predominant extracellular protein expressed when laminarin is used as the sole carbon source. The protein folds into a curved β-sandwich homologous to those of other known GH16 enzyme structures (especially κ-carrageenase from *Pseudoalteromonas carrageenovora* and β-agarase from *Zobelia galactanivorans*). A notable likeness is also evident with the related glycoside hydrolase family 7 (GH7) enzymes. A mammalian lectin, p58/ERGIC, as well as polysaccharide lyase (PL7) enzymes also showed significant similarity to Lam16A. The enzyme has two potential N-glycosylation sites. One such site, at Asn43, displayed a branched heptasaccharide sufficiently stabilized to be interpreted from the X-ray diffraction data. The other N-glycosylation motif was found close to the catalytic centre and is evidently not glycosylated.

1. Introduction

The basidiomycete fungus *Phanerochaete chrysosporium* is among the most extensively studied wood-decaying fungi. It is found in forest litter and on fallen trees and is able to degrade all major components of wood, including cellulose, hemicelluloses and lignin (Eriksson *et al.*, 1990). As with other white-rot fungi, the lignin portion of wood is attacked at an early stage, causing a whitening of the decaying wood.

The complete genome of *P. chrysosporium* strain RP78 has been sequenced (Martinez *et al.*, 2004) and the genome assembly and gene models have recently been updated (Wymelenberg *et al.*, 2006). The current v.2.1 release of the genome database (http://genome.jgi-psf.org/Phchr1/Phchr1.home.html) contains 10 048 gene models, of which 769 were predicted to code for secreted proteins (Wymelenberg *et al.*, 2006). The genome reveals an impressive repertoire of genes for putative biomass-degrading enzymes, with, for example, at least 87 genes encoding putative glycoside hydrolase (GH) enzymes. Peptide analysis by mass spectrometry has verified the presence of the products from 56 different genes in culture filtrate from *P. chrysosporium* grown on cellulose and from 40 genes under standard ligninolytic conditions; 11 of these genes were expressed under both conditions (Wymelenberg *et al.*, 2005, 2006).

Several extracellular oxidative enzymes from *P. chrysosporium* implicated in lignin degradation have been identified and characterized (Cullen & Kersten, 2004) and at least six

enzymes with cellulolytic activity: Cel5A, Cel5B, Cel6A, Cel7C, Cel7D and Cel12A (previously known as EG42, EG38, CBH50, CBH62, CBH58 and EG28, respectively; Henriksson *et al.*, 1999; Uzcategui, Johansson *et al.*, 1991; Uzcategui, Ruiz *et al.*, 1991). A major extracellular β-glucosidase (BGL) of family GH3 has been thoroughly characterized and has been shown to be active primarily on β-1,3-glucosides (Igarashi *et al.*, 2003). Less is known about *P. chrysosporium*'s enzymes for degradation and metabolism of other components.

When *P. chrysosporium* was cultivated with laminarin as the sole carbon source, the predominant extracellular protein was a 36 kDa β-1,3-glucanase (Kawai *et al.*, 2006). The corresponding cDNA was cloned and recombinantly expressed in *Pichia pastoris* for enzymatic characterization. The enzyme consists of a single GH family 16 catalytic module of 298 amino acids and was designated Lam16A. It displays typical endo-1,3(4)-β-glucanase activity (EC 3.2.1.6) with broad substrate specificity and randomly hydrolyzes linear β-1,3-glucan, branched β-1,3/1,6-glucan and β-1,3-1,4-glucan (*e.g.* lichenan). Product analysis suggested that the enzyme requires β-1,3-linked nonsubstituted glucose residues at subsites -2 and -1, whereas it permits a 6-*O*-glucosyl substitution at subsite +1 and both β-1,3 and β-1,4 linkages at the catalytic centre (Kawai *et al.*, 2006).

The physiological role of Lam16A in *P. chrysosporium* is not known, *i.e.* whether its main role is to degrade glucans in wood or other plant biomass for feeding or whether it is mainly involved in metabolism of endogenous glucans. β-1,3-Glucan is present in wood, for example in the form of callose. It is also the main constituent of fungal cell walls, with varying degrees of β-1,6 branching depending on species and developmental stage (Seviour *et al.*, 1992). Furthermore, *P. chrysosporium* can produce extracellular β-1,3-glucan that forms a gel-like sheath suggested to be involved in the attachment of hyphae to the plant cell wall (Pitson *et al.*, 1993; Ruel & Joseleau, 1991; Bes *et al.*, 1987; Sietsma & Wessels, 1981).

GH family 16 contains representatives from all kingdoms, archae, bacteria and eukaryota, and members are found in a diverse range of organisms including fungi, plants, insects, crustaceans and nematodes [Carbohydrate Active Enzymes Database (http://www.cazy.org/; Coutinho & Henrissat, 1999)]. Based on substrate specificity, phylogenetic analysis and conserved structural features (Allouch *et al.*, 2003; Michel *et al.*, 2001; Strohmeier *et al.*, 2004) at least five subgroups can be distinguished: (i) κ-carrageenases/1,4-β-galactanases, (ii) agarases/1,4-β-galactanases, (iii) nonspecific 1,3(4)-β-glucanases, (iv) lichenases/1,3-1,4-β-D-glucan endohydrolases and (v) xyloglucan transglucosylases/hydrolases (XTHs).

Three-dimensional structures have been reported in all subgroups except subgroup (iii), *e.g.* (i) *Pseudoalteromonas carrageenovora* κ-carrageenase (Michel *et al.*, 2001), (ii) *Zobellia galactinovorans* β-agarases A and B (Allouch *et al.*, 2003), (iv) *Bacillus licheniformis* 1,3-1,4-β-glucanase (Hahn *et al.*, 1995) and (v) *Populus tremula* $\times$ *tremuloides* Xet16A (Johansson *et al.*, 2004). They have a common curved β-sandwich fold made up of two antiparallel β-sheets, similar to the folds of GH families 7, 11 and 12.

Enzymes in these families and in GH16 have a retaining reaction mechanism with two consecutive inverting steps that results in a net retention of the β-anomeric configuration of the reactive sugar unit (Koshland, 1953; Sinnott, 1990). In the first step, one carboxylic acid residue, the catalytic acid/base, protonates the glycosidic oxygen, while another carboxylate residue, the catalytic nucleophile, attacks and forms a covalent bond with the anomeric carbon when the glycosidic bond breaks. In the second step, the covalent bond of this glycosyl-enzyme intermediate is hydrolysed by a water molecule that is activated by the catalytic acid/base.

The families GH16 and GH7 have been grouped together in clan GH-B. They have similar catalytic sites and have probably evolved from a common ancestor (Michel *et al.*, 2001). In GH7 and subgroups (i), (ii) and (iii) of GH16 the catalytic motif E*X*D*XX*E contains a β-bulge, whereas the motif E*X*D*X*E in GH16 subgroups (iv) and (v) is one residue shorter and forms a regular β-strand.

In the *P. chrysosporium* genome there are at least 20 putative GH16-like genes (http://genome.jgi-psf.org/Phchr1/Phchr1.home.html). So far, the expression of only two of these genes has been experimentally verified. Lam16A corresponds to gene model 10833 in v.2.1 (pc.78.37.1 in v.1.0), while peptides from model 123909 (pc.22.125.1 in v.1.0), which shows 77% amino-acid sequence identity with Lam16A, were identified in culture filtrate from cellulose cultures (Wymelenberg *et al.*, 2005).

In this study, we present the three-dimensional apo structure of *P. chrysosporium* Lam16A, which is the first structure representative from subgroup (iii). The structure, refined at 1.3 Å resolution, was solved by X-ray crystallography using native S atoms in the protein for phasing.

2. Materials and methods

2.1. Protein preparation and crystallization

Wild-type *P. chrysosporium* Lam16A was heterologously expressed in *Pichia pastoris* and purified as described in previous work (Kawai *et al.*, 2003, 2006). The protein was concentrated to approximately 25 mg ml^{-1} prior to the crystallization experiments using a Vivaspin protein-concentration membrane with a molecular-weight cutoff of 10 kDa. The protein concentration was determined by measuring the absorbance at 280 nm and using a calculated extinction coefficient of 63 900 M^{-1} cm^{-1}.

The protein used for the crystallization experiments was in a buffer solution of 20 m*M* potassium phosphate pH 7.0, 80 m*M* KCl and 0.02% sodium azide. Initial crystallization conditions for Lam16A were obtained using the JCSG+ Core96 screen (Newman *et al.*, 2005). Crystals were obtained in several of the conditions in the screen.

The final condition that gave rise to the best diffracting crystal was a crystallization agent containing 20%(*w*/*w*) PEG 3350 (polyethylene glycol, average size 3350 g mol^{-1}) and 0.2 *M* ammonium nitrate at 293 K using the sitting-drop vapour-diffusion method (McPherson, 1982). Crystallization

Table 1
Data-collection and phasing statistics.

Values in parentheses are for the highest resolution shell.

	SAD	Native
Data collection		
Beamline	ID14-4	MAX-lab 711
Distance (mm)	80	75/220
Wavelength (Å)	1.775	1.081
No. of images	999	360/60
Angle of total revolution (°)	499.5	180/180
Oscillation (°)	0.5	0.5/3
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 38.2, b = 47.0, c = 152.2, $\alpha = \beta = \gamma$ = 90	a = 38.3, b = 47.4, c = 152.2, $\alpha = \beta = \gamma$ = 90
Resolution (Å)	47.0–2.00 (cutoff)	76.5–1.34 (merged)
Unique reflections	19367 (2739)	54024 (5635)
Redundancy†	19.0 (18.8)	6.5 (6.1)
Completeness‡ (%)	99.6 (98.8)	85.6 (64.3)
Anomalous completeness (%)	99.8 (99.2)	—
$R_{\rm merge}$ (%)	4.8 (9.9)	5.5 (12.8)
$I/\sigma(I)$	53.7 (27.5)	24.8 (13.3)
Phasing		
Resolution cutoff (Å)	2.5	
No. of sites (S atoms)	13	
Overall phasing power	1.081	
Overall figure of merit		
Acentric reflections	0.37	
Centric reflections	0.11	

† Multiplicity. ‡ Not anomalous.

drops were prepared by mixing equal amounts of protein solution (25 mg ml^{-1}) and crystallization agent to a final drop size of 2 µl. Large single crystals grew to a maximum size of 0.5 mm in all directions within 2–3 d of seeding the crystallization drops with crystallization nuclei from previously obtained Lam16A crystals. A horse hair was used to transfer crystallization nuclei; the hair was dipped into a drop from an earlier batch (where small crystals were already visible) and swiped through a new drop (which had been allowed to equilibrate for a couple of days beforehand).

2.2. X-ray data collection

The initial Lam16A data set to 2.0 Å resolution, which was used to solve the structure by sulfur SAD, was collected on beamline BM14 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France at a wavelength of 1.775 Å (6.984 keV) from a single crystal at 100 K using 35% PEG 3350 as cryoprotectant. The crystal used for phasing was obtained in the presence of 0.6 m*M* Baker's dimercurial (1,4-diacetoxymercuri-2,3-dimethoxybutane) and was initially intended as a heavy-atom derivative. However, the crystal did not show significant anomalous signal from mercury and was instead used for sulfur SAD data collection. A 1.34 Å resolution native wild-type Lam16A data set, which was used for final structure refinement, was collected at beamline ID711 at MAX-lab, Lund, Sweden from a single crystal at 100 K. The X-ray data sets were processed using *MOSFLM* (Leslie, 1992). The *CCP*4 package (Collaborative Computational Project, Number 4, 1994) was used for subsequent scaling with *SCALA* and other data processing. A set representing 5.0% of the total reflections were set aside and used to monitor $R_{\rm free}$. Data-collection and processing statistics are summarized in Table 1.

Table 2
Refinement statistics.

Resolution (Å)	45.3–1.35
Unique reflections in set	51211
Reflections in test set	2739
$R_{\rm work}$ (%)	14.7
$R_{\rm free}$ (%)	16.8
R.m.s.d. for bond distances (Å)†	0.010
R.m.s.d. for bond angles (°)†	1.4
No. of amino-acid residues ($B_{\rm ave}$, Å^2)	298 (8.03)
No. of waters molecules ($B_{\rm ave}$, Å^2)	388 (20.33)
No. of sugar residues ($B_{\rm ave}$, Å^2)	7 (23.59)
Ramachandran outliers‡ (%)	2.2

† Root-mean-square distance from the ideal values of Engh & Huber (1991). ‡ Ramachandran plot generated by *MOLEMAN*2 (Kleywegt & Jones, 1996).

2.3. Structure solution and refinement

Initial positions for all 13 naturally occurring S atoms of Lam16A could readily be identified using *SHELXC* and *SHELXD* (Schneider & Sheldrick, 2002). Subsequent heavy-atom refinement, density modification and initial structure modelling was performed using *autoSHARP* (de La Fortelle & Bricogne, 1997). The sulfur SAD phases obtained from *autoSHARP* produced an electron-density map which was traced using *ARP*/*wARP* (Collaborative Computational Project, Number 4, 1994).

The structure was refined with alternating cycles of model building using *O* (Jones *et al.*, 1991) and maximum-likelihood refinement using *REFMAC*5.0 (Murshudov *et al.*, 1997; Collaborative Computational Project, Number 4, 1994). Most water molecules in the structure models were located automatically using the water-picking protocols in the refinement program and were then manually selected or discarded by visual inspection. A summary of refinement statistics is given in Table 2.

Structural comparisons were made with *O*. The *Beta-Spider* algorithm was used to determine the location and extent of β-strands (Parisien & Major, 2005). Figures were prepared with *PyMOL* (DeLano, 2002), except for Figs. 3 and 4, which was prepared with *O*, *MOLRAY* (Harris & Jones, 2001) and *PovRay* (http://www.povray.org/). The coordinates for the final model and the structure-factor amplitudes have been deposited in the Protein Data Bank (Bernstein *et al.*, 1977) and have been assigned access code 2cl2. The final electron-density maps for the structure are available for viewing as part of the Electron Density Server (EDS) service at http://eds.bmc.uu.se/eds/ (Kleywegt *et al.*, 2004).

3. Results and discussion

3.1. Crystallization, structure solution and quality of the final model

P. chrysosporium laminarinase Lam16A was expressed in *Pichia pastoris* and crystallized in space group $P2_12_12_1$, with unit-cell parameters a = 38, b = 47, c = 152 Å, a calculated $V_{\rm M}$

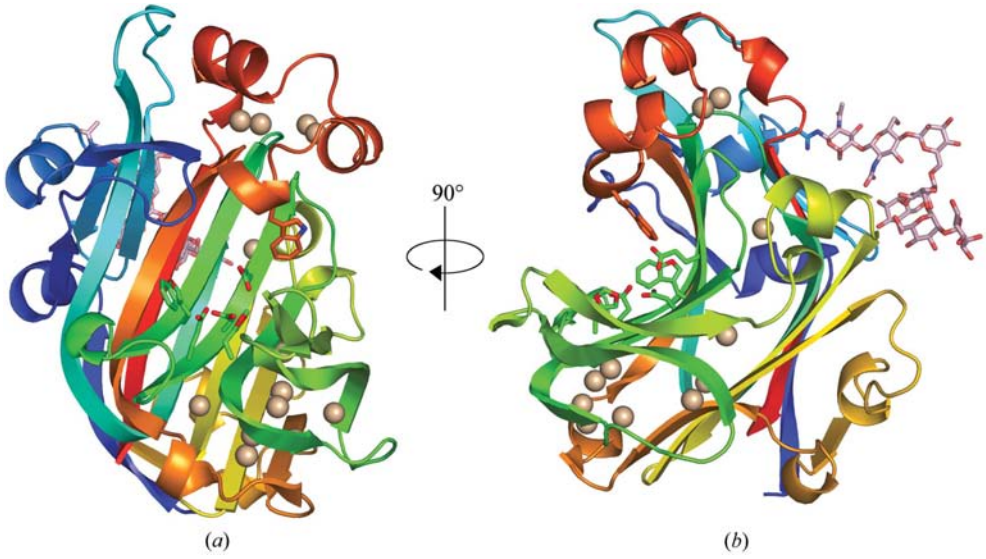

Figure 1
Secondary-structure representation of laminarinase Lam16A from *P. chrysosporium*, rainbow coloured blue to red from the N-terminus to the C-terminus. Side chains are shown for characteristic residues at the catalytic centre: the proposed nucleophile Glu115, the acid/base Glu120 and the assisting aspartate Asp117, the histidine His133 that flanks the nucleophile and the tryptophans Trp103 and Trp257 at subsites −1 and +1, respectively. The glycosylation at Asn43 is shown with pink C atoms. The beige spheres depict all the protein's native S atoms (from methionine and cysteine). β-Sheets were assigned using *Beta-Spider* (Parisien & Major, 2005).

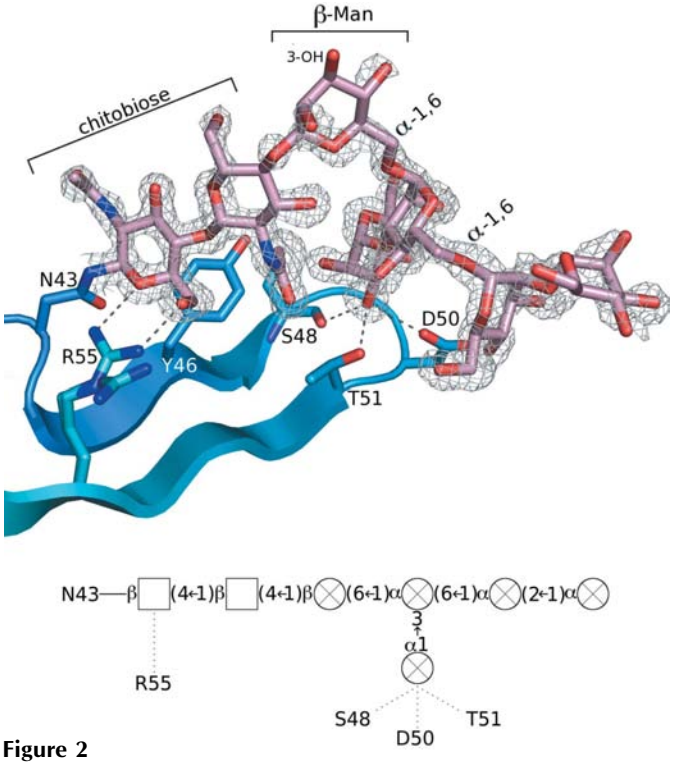

Figure 2
The N-glycosylation on Lam16A. Seven sugar residues of the attached *N*-glycan at Asn43 could be unambiguously positioned in the electron-density map ($2F_o - F_c$, contoured at σ = 1.0 Å around the carbohydrate model). The chitobiose moiety and the α-1,6-arm of the β-mannose residue are well ordered and make several interactions with protein residues, while there is no clear density for the 3-hydroxyl of the β-mannose or any attached α-1,3 arm at this position. In the schematic representation, GlcNAc is represented by squares and Man by circles containing crosses, respectively. Putative hydrogen bonds are shown as dashed lines.

of 1.9 Å^3 Da^{-1} (Matthews, 1968) and a solvent content of 35% with one protein molecule (MW = 36 kDa) in the asymmetric unit. Initial crystallization screening was performed with enzymatically deglycosylated Lam16A, but similar crystals could also be obtained with the glycosylated form, which was then used for further experiments. The structure could not be solved by molecular replacement using known GH16 structures as search models, but was instead solved by the SAD (single-wavelength anomalous dispersion) method using the enzyme's native S atoms for phasing. The SAD diffraction data were collected from a crystal obtained by cocrystallization with Baker's dimercurial and were first used in an attempt at phasing using mercury. The crystal had thus been subjected to both mercury exposure and potentially damaging X-ray radiation prior to the collection of the SAD data set. Nevertheless, the quality of the data was sufficient to solve the structure. A highly redundant (average multiplicity of 19) data set with high anomalous completeness (99%) was obtained by collecting 500° of crystal rotation. The diffraction data were cut at 2.5 Å resolution to determine the sulfur substructure using *SHELXD*. All 13 S-atom sites were found, corresponding to the four disulfide bonds and five methionines present in the protein. Their positions in the structure are shown in Fig. 1. *ARP*/*wARP* then automatically fit 290 of the 298 amino-acid residues (R_{work}/R_{free} = 0.20/0.30) into the density. Resolution was extended to 1.34 Å by merging the SAD data set with diffraction data collected from a crystal without heavy-atom addition (Table 1). The structure refinement yielded final R_{work} and R_{free} values of 0.15 and 0.17, respectively, and all residue side chains could be positioned in the final electron-density maps. Data-collection, phasing and refinement statistics are provided in Tables 1 and 2. The final model contains 298 amino-acid residues, 388 water molecules and seven carbohydrate residues bound to Asn43 (Fig. 2). There are two *cis*-prolines, Pro111 and Pro127. Alternate conformations have been modelled for residues Asp35, Arg55 and Ser282. The Ramachandran plot generated by *MOLEMAN*2 (Kleywegt & Jones, 1996) identified 2.2% of the non-glycine residues as φ, ψ outliers (Thr81, Thr151, Cys236, Asp243, Cys254 and Trp257).

3.2. Overall structure

Antiparallel β-strands form a curved β-sandwich seven strands wide in the concave/inner sheet and six strands wide in

the convex/outer sheet of the sandwich. There are also ten single-turn or double-turn α-helices interspersed on the periphery of the structure (Fig. 1). The protein forms an irregular ellipsoid with approximate dimensions 60 × 40 × 30 Å. The four disulfide bonds are distributed pairwise on either side of the cleft. Cys138–Cys236 and Cys155–Cys165 (Fig. 1) seem to stabilize the three twisted β-strands which appear almost perpendicular to the concave (cleft-side) β-sheet. The other disulfide bonds, Cys96–Cys269 and Cys254–Cys273, stabilize loops on the opposite edge of the cavity (Fig. 1). The two *cis*-prolines, Pro111 and Pro127, are separated by a length of β-strand containing the three catalytic residues.

The concave β-sheet and connecting regions at the periphery combine to form a cleft approximately 30 Å in length, 12 Å deep and 8 Å wide which cuts across the middle of the enzyme. The aromatic residues lining this cleft are thought to participate in binding of the polysaccharide substrate. The cleft is sufficiently long to accommodate at least six sugar residues of a β-1,3-glucan substrate, with three subsites on either side of the catalytic centre, in accordance with recent enzymatic characterization (Kawai *et al.*, 2006).

The presumed catalytic residues, Glu115 (nucleophile), Asp117 and Glu120 (acid/base), are located on the same β-strand near the middle of the cleft (Fig. 1), with a characteristic β-bulge between Asp117 and Glu120. These residues superimpose neatly onto corresponding residues of other GH family 16 enzymes (see Fig. 5).

3.3. N-Glycosylation

There are two potential N-glycosylation sites, N-*X*-S/T, in the Lam16A sequence at Asn43 and Asn250. Asn43 is clearly glycosylated, with sufficient electron density for modelling of no less than seven sugar residues: two *N*-acetylglucosamines (GlcNac or NAG), the β-mannose unit and four residues of the Man(α1-6) arm (Fig. 2). The β-mannose unit is farthest from the protein surface and is less well defined. There is no density for its 3-hydroxyl or any mannose residues of the Man(α1-3) arm probably attached at this position. In contrast, the Man(α1-6) arm is visible and branches again. The subsequent Man(α1-3) arm consists of a single mannose residue whose 3-hydroxyl group is 2.9 Å from the side-chain O atoms of Thr51 and Asp50 and 3.0 Å from the carbonyl O atom of Ser48 and is apparently stabilized thus by hydrogen bonds. One of the conformations of the double conformer Arg55 is also within hydrogen-bonding distance of the first NAG, with the side-chain N atoms 2.9 Å from O6 and 3.4 Å from the ring O atom.

The other potential N-glycosylation site is clearly not glycosylated. Asn250 is located on a strand adjacent to the strand containing the Trp103 platform residue and two strands from the catalytic centre. Its side chain is positioned in the binding cleft close to the catalytic centre and the amide forms hydrogen bonds to Trp103 N^{ε} (O^{δ}, 2.8 Å) and Tyr33 OH ($N^{\delta 2}$, 3.1 Å). There is a cavity above Asn250 which could potentially accommodate an attached GlcNAc unit, but this cavity is occupied by discrete water molecules. Their electron density (and also that of surrounding protein atoms) is well defined. Owing to the limited space, a carbohydrate at this position would yield distinct electron density if present. Furthermore, any glycosylation here would occupy at least part of the −1 and −2 subsites and effectively block substrate binding to the enzyme, thus resulting in an inactive enzyme.

The crystallographic evidence clearly negates glycosylation at Asn250. It is nevertheless interesting that the protein has maintained an N-glycosylation motif at such a critical point. Apparently, there must be other sequence features that prevent glycosylation at this site.

3.4. Comparison with related enzyme structures

Previous comparative studies have divided the family 16 glycoside hydrolases into five different subfamilies: (i) κ-carrageenases, (ii) β-agarases, (iii) nonspecific 1,3(4)-β-glucanases, (iv) 1,3-1,4-β-glucanases and (v) xyloglucan transglucosylases/hydrolases (XTHs) (Strohmeier *et al.*, 2004; Allouch *et al.*, 2003). These subfamilies can be grouped by the presence (i, ii, iii) or absence (iv, v) of a β-bulge at the catalytic site.

Structural comparisons using *DALI* (Holm & Sander, 1996) and subsequent least-squares alignment of the listed proteins using *O* (Jones *et al.*, 1991) verify that Lam16A is highly similar to other GH16 proteins (Table 3). Enzymes from the κ-carrageenase (i) and β-agarase (ii) subfamilies were structurally more similar to the studied laminarinase (iii) template, as judged by somewhat higher *DALI Z* scores, larger number of identical amino acids and by the preservation of the β-bulge at the catalytic centre. However, the r.m.s.d. values are not significantly higher for the representatives of the 1,3-1,4-β-glucanase (iv) and xyloglucan endotransglycosylase (v) subfamilies (Table 3).

Cysteine residues and disulfide bonds are not at all conserved between Lam16A and other solved GH16 structures and the structural homology is limited to the core of the β-sandwich. Many loop regions that connect the β-strands of the sandwich are quite different in length and structure (Fig. 3). Features that are characteristic of Lam16A include a short α-helix in segment 35–43 (upper left of Fig. 1*a* and Fig. 3). Region 137–155 bends away from the binding cleft and forms an additional seventh β-strand at the edge of the concave β-sheet (lower right of Fig. 1*a* and Fig. 3). There is a long insert at 201–227 that folds onto the convex/outer β-sheet (lower right of Fig. 1*b*). The segment at 257–280 is longer and contains two short α-helices (upper right in Fig. 1*a* and Fig. 3). Finally, the loops 62–68 and 107–114 are shorter than in the subfamily (i) representatives, but not compared with the others.

The substrate-binding cleft of Lam16A (iii) is open along its entire length, as in β-agarase A (ii), 1,3-1,4-β-glucanase (iv) and Xet16A (v), while κ-carrageenase and β-galactosidase (i) have loops that fold over and cover part of the cleft. In Lam16A the cleft forms a rather narrow and straight canyon

Table 3
Laminarinase Lam16A *DALI* structural database alignment (Holm & Sander, 1996) and data from subsequent alignment using *LSQMAN* (Kleywegt, 1996).

Protein	Family†	PDB code	Sequence length	Aligned fragment length‡	Aligned amino acids§	Identical amino acids	R.m.s.d.¶ (Å)	*Z* score††
Ps. carrageenovora κ-carrageenase	GH16 (i)	1dyp	271	257 (39–295)	171	29	1.86	18.1
Z. galactanivorans β-agarase	GH16 (ii)	1o4y	270	249 (37–285)	169	30	1.73	17.7
Z. galactanivorans β-agarase B	GH16 (ii)	1o4z	353	276 (74–349)	170	23	1.76	—
C. perfringens endo-β-galactosidase	GH16 (i)	1ups	404	277 (2–278)	156	29	1.85	15.7
Pop. tremula × *tremuloides* XET 16A	GH16 (v)	1un1	431	390 (31–421)	139	20	2.02	—
Rattus norvegicus P58/ERGIC	—	1gv9	260	225 (50–274)	131	18	1.94	12.4
P. chrysosporium cellobiohydrolase Cel7D	GH7	1gpi	431	390 (31–421)	139	20	2.02	—
Pseudomonas aeruginosa alginate lyase	PL7	1vav	222	216 (6–221)	122	11	2.23	11.3
F. succinogenes 1,3-1,4-β-D-glucanase	GH16 (iv)	1mve	243	226 (4–229)	176	27	1.80	11.1
B. licheniformis β-glucanase	GH16 (iv)	1gbg	214	199 (5–213)	160	21	1.64	—
Corynebacterium sp. polyguluronate lyase	PL7	1uai	224	210 (14–223)	108	17	2.12	11.0
B. subtilis hypothetical protein	—	1oq1	223	212 (5–216)	121	8	2.17	10.8
B. subtilis β-1,4-xylosidase	GH43	1yif	533	205 (329–533)	113	4	2.37	10.7
Alteromonas sp. 272 alginate lyase	PL18	1j1t	233	212 (18–229)	113	14	2.29	10.5
Griffonia simplicifolia lectin	—	1led	243	232 (5–236)	127	10	2.26	10.2
Canis familiaris calnexin	—	1jhn	424	371 (72–442)	117	10	2.33	9.8
Vibrio cholerae neuraminidase	CBM40	1kit	757	188 (26–213)	108	4	2.04	9.7
Trypanosoma rangeli sialidase	GH33	1mz5	638	192 (419–610)	113	12	2.03	9.3
Shewanella oneidensis hypothetical protein	—	2a5z	262	204 (50–253)	110	4	1.98	9.1
Aspergillus kawachii α-L-arabinofuranosidase	CBM42	1wd3	482	286 (28–313)	100	4	2.22	8.7
Homo sapiens thromospondin	—	1ux6	350	210 (940–1149)	96	4	2.15	8.5
C. tetani tetanus toxin C-fragment	—	1a8d	452	210 (21–230)	109	12	2.40	8.4

† Family nomenclature as in CAZy (Coutinho & Henrissat, 1999); subfamilies are denoted in parentheses. ‡ Length of contiguous superimposing fragment; the start and end of this fragment are denoted within parentheses. § Count of all C^{α} alignments within the r.m.s.d. = 3.8 Å cutoff limit. ¶ Root-mean-square distance of C^{α} atoms in the least-squares superimposition of the aligned C^{α} atoms. †† *Z* score calculated by *DALI* (Holm & Sander, 1996).

(Fig. 4), whereas it is somewhat wider in β-agarase A and 1,3-1,4-β-glucanase and much wider in Xet16A.

As shown in Fig. 5, the active sites are very similar in clan GH-B (GH16 and GH7). The positions of the catalytic residue side chains are virtually identical, despite the presence of a β-bulge in some members. All glycoside hydrolases have a hydrophobic platform at the bottom of subsite −1, usually an aromatic residue (Nerinckx *et al.*, 2003). This residue, Trp103 in Lam16A, tends to be a tryptophan for subfamilies (i), (ii) and (iii), a phenylalanine for 1,3-1,4-β-glucanases (iv) and a tyrosine for xyloglucan endotransferases and GH7 enzymes (v). On the other side, in subsite +1 the nucleophile Glu115 is flanked by a histidine residue in Lam16A, κ-carrageenase, β-agarase and *P. chrysosporium* Cel7D (Muñoz *et al.*, 2001), while the corresponding residue is phenylalanine in Xet16A and isoleucine in *Fibrobacter succinogenes* 1,3-1,4-β-D-glucanase. At site +1 there is a potential tryptophan sugar-binding platform (Trp257 in Lam16A) in all structures except *Clostridium perfringens* endo-β-galactosidase. It has the highest variation among the studied residues. However, despite the different positions of the C^{α} atom, the tryptophan side chains tend to overlap.

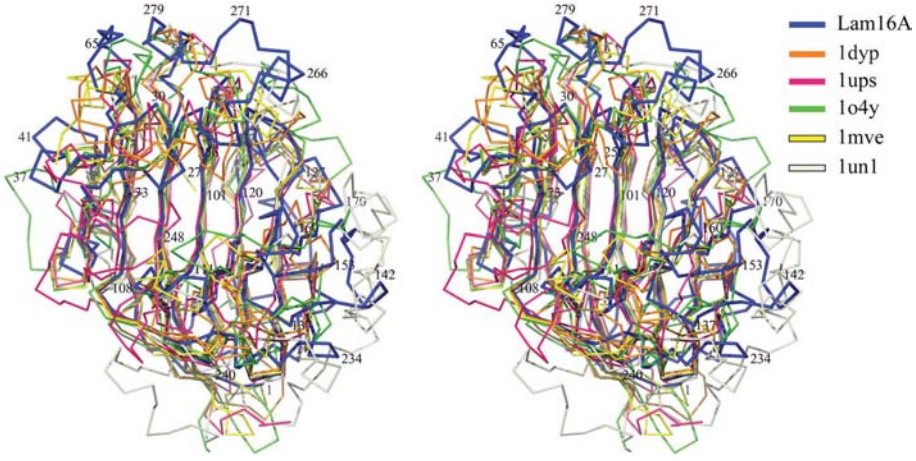

Figure 3
Stereoview of superimposed C^{α} traces of GH16 enzyme structures. *P. chrysosporium* Lam16A of subfamily (iii) is in blue with selected residue numbers indicated. Subfamily (i) is represented by *Ps. carrageenovora* κ-carrageenase (PDB code 1dyp; orange) and *C. perfringens* endo-β-galactosidase (1ups; red; only the GH16 domain is shown), subfamily (ii) by *Z. galactanivorans* β-agarase A (1o4y; green), subfamily (iv) by *F. succinogenes* 1,3-1,4-β-D-glucanase (1mve; yellow) and subfamily (v) by *Pop. tremula* × *tremuloides* xyloglucan endotransglycosylase Xet16A (1un1; light beige).

At present, crystallographic data on carbohydrate binding in GH16 enzymes is limited to subfamilies (iv) and (v). Among the lichenases (iv), three structures have been published with oligosaccharides bound at the 'non-reducing' side of the cleft, *i.e.* subsites −1 to −4: the hybrid B*acillus amyloliquefaciens*/*macerans* 1,3-1,4-β-glucanase H(A16-M) with a covalently bound epoxybutyl-cellobioside inhibitor (Keitel *et al.*, 1993), an inactive mutant of the same enzyme in complex with a β-glucan tetrasaccharide (Gaiser *et al.*, 2006) and *F. succinogenes* 1,3-1,4-β-glucanase in complex with β-1,3-1,4-

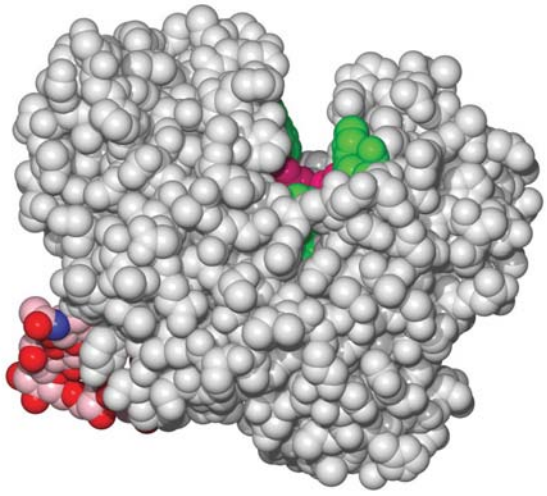

Figure 4
Space-filling model of Lam16A. The proposed catalytic residues (magenta) are located in the middle of a substrate-binding cleft that forms a straight and rather narrow canyon. There are three tryptophan residues in the cleft (green): Trp110 at proposed subsite −2, Trp103 at −1 and Trp257 at +1. The attached *N*-glycan is shown with pink C atoms.

cellotriose (Tsai *et al.*, 2005). In subfamily (v) the structure of *Pop. tremula* × *tremuloides* Xet16A in complex with a xyloglucan nonasaccharide illustrates binding at the 'reducing' side of the cleft, in subsites +1 to +3 (Johansson *et al.*, 2004). Superposition of these complex structures with Lam16A reveals that apart from the conserved residues around the catalytic centre mentioned above, no other carbohydrate-interacting residues are preserved in similar positions in the cleft of Lam16A. Furthermore, in subsites −2 and −3 the side chains of Arg73 and Glu107 extend into the cleft and overlap with glucosyl hydroxyls in the lichenase complexes (O6 in site −2 and O2 in site −3). It thus seems that β-glucan substrates will bind quite differently in Lam16A. However, it is difficult at this stage to draw further conclusions about substrate binding and specificity from a structural point of view.

The structural homology search (Table 3) revealed GH7 and PL7 (Yamasaki *et al.*, 2004) enzymes and a mammalian lectin p58/ERGIC-53 (Velloso *et al.*, 2002) to have structures similar to that of Lam16A. Of particular interest was p58/ERGIC-53, which may be a control mechanism for correct mannosylation in the Golgi apparatus (Schrag *et al.*, 2003;

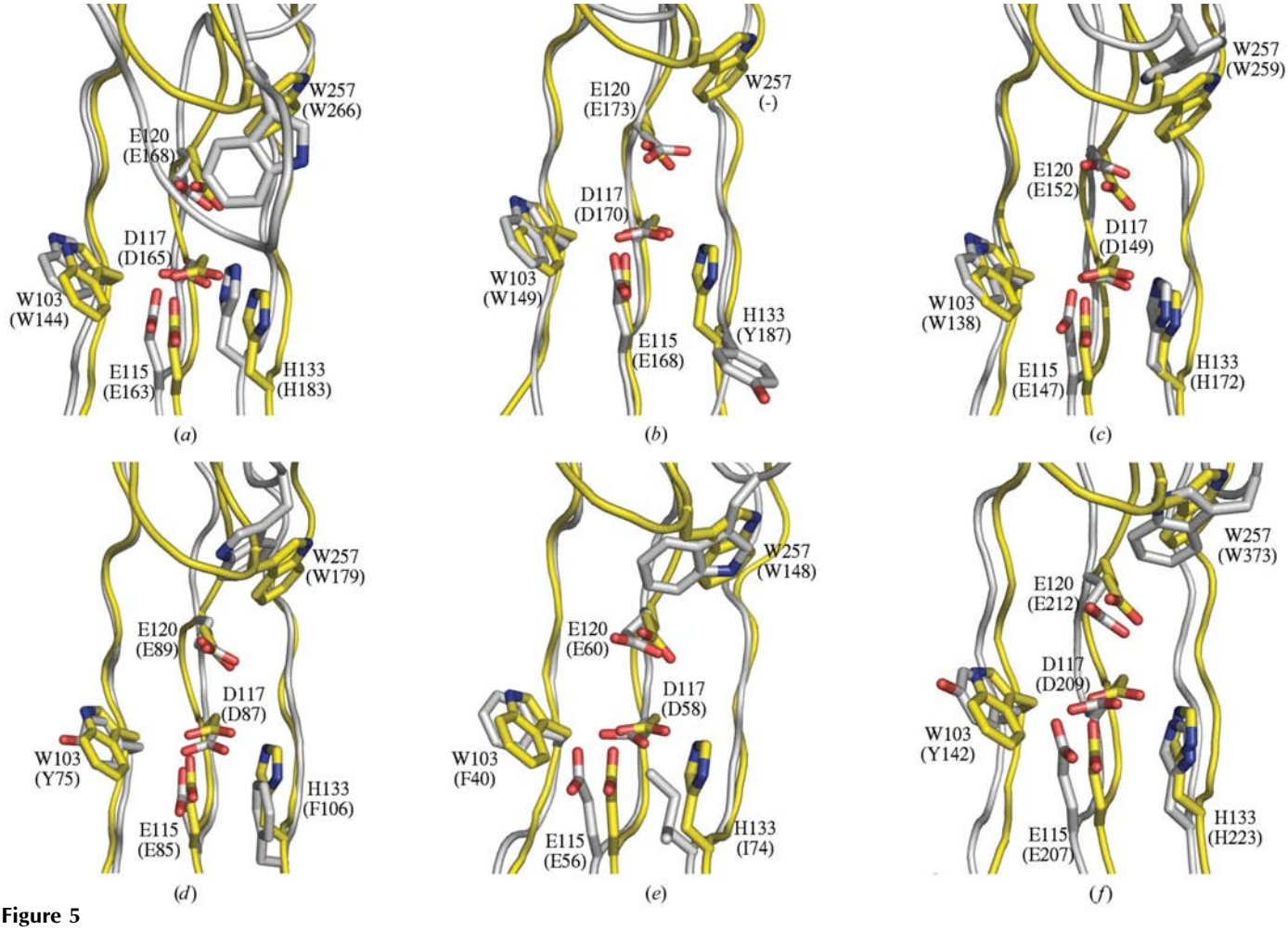

Figure 5
Superposition showing the active site of Lam16A (yellow) superposed onto those of related enzymes (white, with amino-acid labels in parentheses): (*a*) *Ps. carrageenovora* κ-carrageenase (PDB code 1dyp), (*b*) *C. perfringens* endo-β-galactosidase (1ups), (*c*) *Z. galactanivorans* β-agarase A (1o4y), (*d*) *Pop. tremula* × *tremuloides* xyloglucan endotransglycosylase Xet16A (1un1), (*e*) *F. succinogenes* 1,3-1,4-β-D-glucanase (1mve), (*f*) *P. chrysosporium* cellobiohydrolase Cel7D (1gpi).

Sacchettini *et al.*, 2001). Further down the list, in addition to other glycosyl hydrolases, there are polysaccharide lyases (PL7, PL18) as well as potentially interesting cell-adhesion proteins from mammals and microbes. As structural data accumulate, structural homologies between such proteins and glycoside hydrolases (such as the laminarinase described here) may help to elucidate the carbohydrate-binding behaviour of structurally similar yet mechanistically unrelated proteins.

We would like to thank Jean-Baptiste Reiser, Martin Walsh and Hassan Belrhali at beamline BM14, ESRF and Yngve Cerenius at beamline I711, MAX-lab for support during data collection. Financial support from the Knut and Alice Wallenberg Foundation through the Swedish Center for Tree Functional Genomics is gratefully acknowledged. This research was also partly supported by the Japan Society for the Promotion of Science (JSPS) through a Grant-in-Aid for Scientific Reseach to M. Samejima (No. 17380102) and a Research Fellowship to RK (No. 11536).

References

Allouch, J., Jam, M., Helbert, W., Barbeyron, T., Kloareg, B., Henrissat, B. & Czjzek, M. (2003). *J. Biol. Chem.* **278**, 47171–47180.
Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Bes, K., Pettersson, B., Lennholm, H., Iversen, T. & Eriksson, K. (1987). *Biotechnol. Appl. Biochem.* **9**, 310–318.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Coutinho, P. & Henrissat, B. (1999). *Recent Advances in Carbohydrate Bioengineering*, pp. 3–12. Cambridge: The Royal Society of Chemistry.
Cullen, D. & Kersten, P. (2004). *The Mycota*, Vol. 3, edited by R. Brambl & G. A. Marzluf, pp. 249–273. Berlin: Springer–Verlag.
DeLano, W. L. (2002). *The PyMOL User's Manual.* San Carlos, CA, USA: DeLano Scientific.
Engh, R. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Eriksson, K.-E., Blanchette, R. & Ander, P. (1990). *Microbial and Enzymatic Degradation of Wood and Wood Components.* Berlin: Springer–Verlag.
Gaiser, O., Piotukh, K., Ponnuswamy, M., Planas, A., Borriss, R. & Heinemann, U. (2006). *J. Mol. Biol.* **357**, 1211–1225.
Hahn, M., Pons, J., Planas, A., Querol, E. & Heinemann, U. (1995). *FEBS Lett.* **374**, 221–224.
Harris, M. & Jones, T. A. (2001). *Acta Cryst.* D**57**, 1201–1203.
Henriksson, G., Nutt, A., Henriksson, H., Pettersson, B., Stahlberg, J., Johansson, G. & Pettersson, G. (1999). *Eur. J. Biochem.* **259**, 88–95.
Holm, L. & Sander, C. (1996). *Science*, **273**, 595–603.
Igarashi, K., Tani, T., Kawai, R. & Samejima, M. (2003). *J. Biosci. Bioeng.* **95**, 572–576.
Johansson, P., Brumer, H. R., Baumann, M. J., Kallas, A. M., Henriksson, H., Denman, S. E., Teeri, T. T. & Jones, T. A. (2004). *Plant Cell*, **16**, 874–886.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kawai, R., Igarashi, K., Yoshida, M., Kitaoka, M. & Samejima, M. (2006). *Appl. Microbiol. Biotechnol.* **71**, 898–906.
Kawai, R., Yoshida, M., Tani, T., Igarashi, K., Ohira, T., Nagasawa, H. & Samejima, M. (2003). *Biosci. Biotechnol. Biochem.* **67**, 1–7.
Keitel, T., Simon, O., Borriss, R. & Heinemann, U. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 5287–5291.
Kleywegt, G. J. (1996). *Acta Cryst.* D**52**, 842–857.
Kleywegt, G. J., Harris, M. R., Zou, J.-Y., Taylor, T. C., Wahlby, A. & Jones, T. A. (2004). *Acta Cryst.* D**60**, 2240–2249.
Kleywegt, G. J. & Jones, T. A. (1996). *Structure*, **4**, 1395–1400.
Koshland, D. E. (1953). *Biol. Rev.* **28**, 416–436.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Leslie, A. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
McPherson, A. (1982). *Preparation and Analysis of Protein Crystals.* New York: John Wiley & Sons.
Martinez, D., Larrondo, L. F., Putnam, N., Gelpke, M. D. S., Huang, K., Chapman, J., Helfenbein, K. G., Ramaiya, P., Detter, J. C., Larimer, F., Coutinho, P. M., Henrissat, B., Berka, R., Cullen, D. & Rokhsar, D. (2004). *Nature Biotechnol.* **22**, 695–700.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B. & Dideberg, O. (2001). *Structure*, **9**, 513–525.
Muñoz, I., Ubhayasekera, W., Henriksson, H., Szabo, I., Pettersson, G., Johansson, G., Mowbray, S. & Ståhlberg, J. (2001). *J. Mol. Biol.* **314**, 1097–1111.
Murshudov, G., Vagin, A. & Dodson, E. (1997). *Acta Cryst.* D**53**, 240–255.
Nerinckx, W., Desmet, T. & Claeyssens, M. (2003). *FEBS Lett.* **538**, 1–7.
Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* D**61**, 1426–1431.
Parisien, M. & Major, F. (2005). *Proteins*, **61**, 545–558.
Pitson, S. M., Seviour, R. J. & McDougall, B. M. (1993). *Enzyme Microb. Technol.* **15**, 178–192.
Ruel, K. & Joseleau, J. (1991). *Appl. Environ. Microbiol.* **57**, 374–384.
Sacchettini, J. C., Baum, L. G. & Brewer, C. F. (2001). *Biochemistry*, **40**, 3009–3015.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Schrag, J. D., Procopio, D. O., Cygler, M., Thomas, D. Y. & Bergeron, J. J. M. (2003). *Trends Biochem Sci.* **28**, 49–57.
Seviour, R., Stasinopoulus, S., Auer, D. & Gibbs, P. (1992). *Crit. Rev. Biotechnol.* **12**, 279–298.
Sietsma, J. H. & Wessels, J. G. (1981). *J. Gen. Microbiol.* **125**, 209–212.
Sinnott, M. L. (1990). *Chem. Rev.* **90**, 1171–1202.
Strohmeier, M., Hrmova, M., Fischer, M., Harvey, A. J., Fincher, G. B. & Pleiss, J. (2004). *Protein Sci.* **13**, 3200–3213.
Tsai, L.-C., Shyur, L.-F., Cheng, Y.-S. & Lee, S.-H. (2005). *J. Mol. Biol.* **354**, 642–651.
Uzcategui, E., Johansson, G., Ek, B. & Pettersson, G. (1991). *J. Biotechnol.* **21**, 143–159.
Uzcategui, E., Ruiz, A., Montesino, R., Johansson, G. & Pettersson, G. (1991). *J. Biotechnol.* **19**, 271–285.
Velloso, L. M., Svensson, K., Lahtinen, U., Schneider, G., Pettersson, R. F. & Lindqvist, Y. (2002). *Acta Cryst.* D**58**, 536–538.
Wymelenberg, A. V., Minges, P., Sabat, G., Martinez, D., Aerts, A., Salamov, A., Grigoriev, I., Shapiro, H., Putnam, N., Belinky, P., Dosoretz, C., Gaskell, J., Kersten, P. & Cullen, D. (2006). *Fungal Genet. Biol.* **43**, 343–356.
Wymelenberg, A. V., Sabat, G., Martinez, D., Rajangam, A. S., Teeri, T. T., Gaskell, J., Kersten, P. J. & Cullen, D. (2005). *J. Biotechnol.* **118**, 17–34.
Yamasaki, M., Moriwaki, S., Miyake, O., Hashimoto, W., Murata, K. & Mikami, B. (2004). *J. Biol. Chem.* **279**, 31863–31872.