Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Endo-β-D-1,4-mannanase from *Chrysonilia sitophila* displays a novel loop arrangement for substrate selectivity

Ana Maria D. Gonçalves,[a] Catarina S. Silva,[a] Tânia I. Madeira,[b] Ricardo Coelho,[a] Daniele de Sanctis,[c] Maria Vitória San Romão[b] and Isabel Bento[a]*

[a]Macromolecular Crystallography Unit, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Avenida da República, 2780-157 Oeiras, Portugal, [b]Physiology of Environmentally Conditioned Microbiota Laboratory, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Avenida da República, 2780-157 Oeiras, Portugal, and [c]Structural Biology Group, European Synchrotron Radiation Facility, 6 Rue Jules Horowitz, 38043 Grenoble CEDEX, France

Correspondence e-mail: bento@itqb.unl.pt

Received 15 June 2012
Accepted 3 August 2012

PDB Reference: endo-β-D-1,4-mannanase, 4awe

The crystal structure of wild-type endo-β-D-1,4-mannanase (EC 3.2.1.78) from the ascomycete *Chrysonilia sitophila* (CsMan5) has been solved at 1.40 Å resolution. The enzyme isolated directly from the source shows mixed activity as both an endo-glucanase and an endo-mannanase. CsMan5 adopts the $(\beta/\alpha)_8$-barrel fold that is well conserved within the GH5 family and has highest sequence and structural homology to the GH5 endo-mannanases. Superimposition with proteins of this family shows a unique structural arrangement of three surface loops of CsMan5 that stretch over the active centre, promoting an altered topography of the binding cleft. The most relevant feature results from the repositioning of a long loop at the extremity of the binding cleft, resulting in a shortened glycone-binding region with two subsites. The other two extended loops flanking the binding groove produce a narrower cleft compared with the wide architecture observed in GH5 homologues. Two aglycone subsites (+1 and +2) are identified and a nonconserved tryptophan (Trp271) at the +1 subsite may offer steric hindrance. Taken together, these findings suggest that the discrimination of mannan substrates is achieved through modified loop length and structure.

1. Introduction

Saprophytic (micro)organisms that use the plant cell wall as a nutrient synthesize extensive repertoires of degradative enzymes reflecting the chemical complexity of the recalcitrant substrate (DeBoy *et al.*, 2008; Martinez *et al.*, 2008). A puzzling feature of plant-cell-wall-degrading enzyme systems is the existence of large numbers of closely related enzymes. This diversity of the secretome reflects the evolutionary response to adaptation and the ability to accommodate subtle differences in the structure of the cell-wall polysaccharides. Equally intriguing is the metabolic synergy among the different enzymatic components of the secretome and their varying levels of expression throughout the process of plant-cell-wall degradation.

Hemicellulose represents one of the major and most abundant sources of renewable organic matter. An important class of hemicellulose that is strongly represented in the cell walls of higher plants is mannan. Mannan is a major matrix polysaccharide in the cell walls of angiosperms and may appear as a linear or branched polysaccharide composed of sugar monomers such as D-mannose, D-galactose and D-glucose (Moreira & Filho, 2008; Tailford *et al.*, 2009). The homopolymer contains β-D-1,4-linked mannose sugar residues, while galactomannans correspond to the heteropolymer with α-1,6-galactosyl moieties dispersed throughout the

backbone. Glucomannans are heteropolymers of β-D-1,4-linked mannosyl and glucosyl sugars that may have α-D-1,6-galactoside attached to the mannoside units. Acetylation at O2 or O3 may occur on all forms of mannans and glucomannans, in an irregular pattern for the latter (Moreira & Filho, 2008; Gilbert *et al.*, 2008).

Owing to their structural complexity, the degradation of mannans and derivatized mannans requires the concerted action of a variety of hydrolytic enzymes. The mannan-degrading enzymes are β-mannanase (EC 3.2.1.78) and β-mannosidase (EC 3.2.1.25). They are classified into glycoside hydrolase (GH) families 5, 26 and 113 (Tailford *et al.*, 2009; Hogg *et al.*, 2003) based on amino-acid sequence similarity (http://www.cazy.org; Cantarel *et al.*, 2009) and are all members of the GH-A clan (Henrissat *et al.*, 1995). Other enzymes involved in the removal of side-chain elements are acetyl mannan esterase (EC 3.1.1.6) and α-galactosidase (EC 3.2.1.22) (Moreira & Filho, 2008). Endo-β-D-1,4-mannanases have been isolated from plants, anaerobic and aerobic fungi, eubacteria and archaea (Park *et al.*, 2011). Note that only GH5, the largest family, includes both mannanases and mannosidases. This family contains both bacterial and eukaryotic endo-β-D-1,4-mannanases (EC 3.2.1.78) as well as endo-glucanases (EC 3.2.1.4) and exo-1,3-glycanases, among others, while the smaller GH26 and GH113 families mostly contain enzymes of prokaryotic origin (Hogg *et al.*, 2003). GH5, GH26 and GH113 family members contain a $(\beta/\alpha)_8$-barrel fold and two characteristic catalytic glutamic acid residues (Hogg *et al.*, 2003).

We have focused on the characterization of secreted enzymes from the saprophytic fungus *Chrysonilia sitophila* (Mont.) (Shear & Dodge, 1927; von Arx, 1981; a telomorph of *Neurospora sitophila* and closely related to *N. crassa*) grown on microcrystalline cellulose (Avicel) at specific fermentation times. The fungus *C. sitophila* has been identified as the dominant colonizing species of cork slabs during stabilization, the immediate step after cork-slab boiling, a procedure in industrial cork processing (Oliveira *et al.*, 2003). Its presence on cork slabs inhibits the development of other moulds and exerts important effects on cork for industrial applications: *C. sitophila* does not induce cork-taint development, which is common with other invasive fungi such as *Penicillium* sp. (Silva Pereira *et al.*, 2000), and promotes a softening of the slabs required for further manipulation without disturbing the mechanical and chemical qualities of the cork itself (Silva Pereira *et al.*, 2000, 2006; Rosa *et al.*, 1990). The secretome of *C. sitophila* has been demonstrated to have interesting properties and characteristics regarding the ability to metabolize the cell-wall constituents of cork, cellulose, lignin and suberin, with the latter two known to be particularly recalcitrant components (Vitorino *et al.*, 2007; Centeno & Calvo, 2001). We are interested in characterizing the secretome of *C. sitophila* with regard to cellulose degradation as a function of substrate and time. The wild-type endo-β-D-1,4-mannanase from the *C. sitophila* secretome has been successfully purified and crystallized. Here, we describe the three-dimensional crystal structure of the enzyme solved at 1.40 Å resolution.

2. Materials and methods

2.1. Fungal growth on solid culture and spore suspensions

The fungus *C. sitophila* belongs to a collection of fungi available at the Instituto de Biologia Experimental e Tecnológica (IBET) and has been deposited at the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Germany under code DSM 16514. The preparation of solid medium inoculates and spore solutions has been described elsewhere (Silva Pereira *et al.*, 2000; Vitorino *et al.*, 2007).

2.2. Inoculation in liquid medium and enzymatic extract processing

Cultures were grown in basal medium M1 [modified from Sternberg (1976) and described by Silva Pereira *et al.* (2000)] with the pH adjusted to 6.0 with 0.1 *M* NaOH. M1 containing microcrystalline cellulose (7.5 g l^{-1}) as the sole carbon source was used for increased selectivity. The final concentration of the inoculate was 10^5 spores per litre. Cultures incubated in the dark at 300 K and 80 rev min^{-1} for 50 h were collected and centrifuged (8600*g* for 30 min at 277 K) in order to collect the supernatant containing the crude enzymatic extract. The supernatant resulting from 1 l culture was filtered with a nylon membrane of 45 µm pore size and subsequently concentrated to a final volume of approximately 3 ml with a 30 kDa cutoff PES membrane (Amicon, USA).

2.3. Enzyme purification

The filtered and concentrated enzymatic extract was loaded onto a Mono Q HR 5/5 ion-exchange column previously equilibrated with 20 m*M* Tris–HCl pH 7.6 using an ÄKTA FPLC (GE Healthcare). Elution was achieved using a linear salt gradient from 0 to 500 m*M* NaCl in 40 column volumes. The eluted samples were analysed by SDS–PAGE for purity assessment and estimation of the respective molecular weight. Fractions were further tested for exo-glycosidase activity by incubation with *p*-nitrophenyl-β-D-glucopyranoside (pNPG; see §2.7). Pure endo-1,4-β-D-mannanase enzyme was eluted at approximately 200 m*M* NaCl and the fractions were pooled and concentrated to 6.5 mg ml^{-1} using a 30 kDa cutoff PES membrane (Amicon, USA). Purification yields were typically low for endo-1,4-β-D-mannanase, ranging from 300 to 450 µg per litre of fermentation broth.

2.4. Mass spectrometry and N-terminal sequencing

Isolated and purified endo-1,4-β-D-mannanase was subjected to trypsin digestion and analysed by MALDI-TOF/TOF mass spectrometry (Applied Biosystems Model 4800), a tandem time-of-flight MS/MS system available at the Mass Spectrometry Laboratory at ITQB, Oeiras, Portugal. This system allows peptide identification and peptide sequencing. Peptides were identified using the *Protein Pilot* software from ABI (Paragon Algorithm) and *GPS* from ABI using an in-house *Mascot* server. Processed data were searched against the *N. crassa* database (Broad Institute; http://www.broad.mit.edu/annotation/genome/neurospora/Home.html). Sequencing of

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

PDB code 4awe. Values in parentheses are for the outer shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 57.39, b = 79.94, c = 83.86
Data-collection statistics	
Wavelength (Å)	0.9762
Resolution (Å)	1.40
No. of unique reflections	75187 (10694)
Multiplicity	4.6 (4.0)
Completeness (%)	99.5 (98.2)
Mean $I/\sigma(I)$	6.5 (2.2)
$R_{\rm merge}$† (%)	11.5 (38.9)
Refinement and model statistics	
Resolution (Å)	1.40
No. of reflections used (working set)	71225
$R_{\rm work}$ (%)	16.1
$R_{\rm free}$ (%)	20.2
No. of residues	381
No. of water molecules	524
R.m.s.d. bond lengths (Å)	0.009
R.m.s.d. angles (°)	1.237
Mean B factor (Å^2)	
Main chain/side chain	12.71/15.06
Solvent	26.41
MolProbity statistics	
Ramachandran	
Most favoured (%)	97.35
Allowed (%)	2.64
Outliers (%)	0.00
Rotamer outliers (%)	1.29
Clashscore	0.32

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of observation i of reflection hkl.

the first nine residues of the N-terminus of the pure enzyme in solution and of enzyme crystals was performed by a stepwise Edman degradation reaction using a Procise 491 HT Protein Sequencer (Applied Biosystems) at the Analytical Services Unit, IBET, Oeiras, Portugal.

2.5. Crystallization and data collection

Initial crystals were obtained after 12 d from screenings with a Cartesian Mini-Bee Crystallization Robot (Genomic Solutions) using the JCSG-plus crystallization screen (Molecular Dimensions) in a condition consisting of 0.2 *M* magnesium chloride, 0.1 *M* bis-Tris pH 5.5, 25%(*w*/*v*) PEG 3350. Crystals were further optimized using the hanging-drop vapour-diffusion method at 293 K by mixing equal volumes of protein and crystallization solution now consisting of 0.1 *M* magnesium chloride, 0.1 *M* sodium acetate pH 4.5, 23%(*w*/*v*) PEG 3350. The protein crystallized in the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 57.39, b = 79.94, c = 83.86 Å. The solvent content was estimated as 44.7% based on a Matthews coefficient of 2.22 Å^3 Da^{-1} for one molecule in the asymmetric unit and a unit-cell volume of 384 628 Å^3 (Winn *et al.*, 2011). Suitable crystals were cryocooled using the crystallization solution with the PEG 3350 precipitant incremented to 35%(*w*/*v*) as a cryoprotectant. Data were collected to 1.40 Å resolution from a single crystal on beamline ID-29 at the European Synchrotron Radiation Facility, Grenoble, France (de Sanctis *et al.*, 2012). Diffraction data were integrated with *iMOSFLM* v.0.5.2 (Battye *et al.*, 2011) and were reduced and scaled with *SCALA* (Evans, 2006) from the *CCP*4 suite (Winn *et al.*, 2011). Data-collection statistics are given in Table 1.

2.6. Structure determination and refinement

The structure was solved using the molecular-replacement program *Phaser* (McCoy *et al.*, 2007) with the structure of *Hypocrea jecorina* (formerly *Trichoderma reesei*) endo-β-D-1,4-mannanase (PDB entry 1qno; Sabini *et al.*, 2000) as the search model. The model of *C. sitophila* endo-β-D-1,4-mannanase was built and refined with *Coot* (Emsley & Cowtan, 2004) and *REFMAC* v.5.4 (Murshudov *et al.*, 2011), respectively. The sequence was corrected and any ambiguity relative to side chains was eliminated based on the sequenced gene amplified from the cDNA library (see below). A randomly selected 5% of observed reflections were kept aside for cross-validation. The final model was analysed and validated with *MolProbity* (Table 1; Chen *et al.*, 2010).

2.7. Enzyme assays

Assay mixtures (total volume of 1 ml) consisting of 100 m*M* acetate buffer pH 5.0, 5 m*M* *p*-nitrophenyl-β-D-glucopyranoside and enzyme solution were incubated at 310 K for 30 min. The reaction was stopped by the addition of 250 m*M* sodium carbonate (1 ml) and the absorbance was measured at 410 nm. Specific activities were calculated using the molar extinction coefficient determined for *p*-nitrophenol under the conditions of the assay. Identical procedures were performed for other *p*-nitrophenyl derivatives. Activity against azo-carob galactomannan (Megazyme) was tested following the supplier's instructions, with substrate at 0.5%(*w*/*v*). Units of activity are as defined by the substrate supplier.

2.8. mRNA purification, cDNA generation and library construction

Samples of the fungus *C. sitophila* were collected after approximately 50 h of fermentation in the presence of Avicel as the sole carbon substrate and were further treated with an Oligotex Direct mRNA Mini Kit (Qiagen) for mRNA extraction following the manufacturer's instructions. Samples were subsequently treated with the RNase-Free DNase Set (Qiagen) for efficient removal of DNA. cDNA generation was performed by RT-PCR with a First Strand cDNA Synthesis Kit (AMV) from Roche. cDNA was used as a template for PCR using degenerate primers based on homology from sequence analysis of *N. crassa* endo-β-D-1,4-mannanase (NCU 08412) and the sequence derived from the *C. sitophila* endo-β-D-1,4-mannanase crystal structure; the forward primer comprised a 5′-CACC overhang allowing incorporation into a pENTR/TEV/D-TOPO vector (Invitrogen). The cloned vector was checked for the desired insert and sequenced.

2.9. Accession numbers

Atomic coordinates of *C. sitophila* endo-β-D-1,4-mannanase (CsMan5) have been deposited in the Protein Data Bank as entry 4awe. The DNA sequence of *C. sitophila*

endo-β-D-1,4-mannanase (CsMan5) corresponding to the mature enzyme has been deposited in the EMBL Nucleotide Archive under accession No. HE856289.

3. Results and discussion

3.1. Identification and characterization of wild-type protein

Previous fermentations using Avicel as the sole carbon source (unpublished results) showed enhanced production of cellulases after 40–50 h of incubation. Following a simple purification protocol, we have isolated an enzyme from the fermentation supernatant that shows both endo-glucanase and exo-glucanase activity as assayed by measuring the release of reduced sugar (results not shown). Such apparent promiscuity led to further tests for exo-glycosidase activity using *p*-nitrophenyl-β-D-glucopyranoside and a series of other *p*-nitrophenyl substrates (Fig. 1). The affinity for *p*-nitrophenyl-β-D-mannopyranoside led to tests for endo-activity using the branched hemicellulose azo-carob galactomannan. For this polysaccharide the enzyme showed a specific activity of 627 U mg^{-1}, a value that is comparable to those reported for other β-mannanases in the literature (Hogg *et al.*, 2003; Zhang *et al.*, 2008; Dilokpimol *et al.*, 2011). Although dual activity against glucan-based and mannan-based polysaccharides is unusual in GH5 enzymes, some exceptions have been reported that show diverse substrate recognition (Park *et al.*, 2011; Chhabra *et al.*, 2002; Wu *et al.*, 2011).

Pure protein, henceforth named CsMan5, was subjected to trypsin digestion followed by MALDI-TOF/TOF mass spectrometry. The highest sequence homology of the 44.5 kDa protein was traced to the product of the *NCU08412* gene (formerly annotated as *NCU11068*) from the *N. crassa* genome. The *NCU08412* gene encodes a 400-residue protein with a calculated molecular mass of 44 974 Da and a theoretical pI of 5.34. The *NCU08412* gene product, initially curated as a conserved hypothetical protein, has more recently been annotated as an endo-β-D-1,4-mannanase (Sun & Glass, 2011). The gene product from *N. crassa* presents a single module corresponding to the catalytic domain; the annotated gene does not include a cellulose-binding domain.

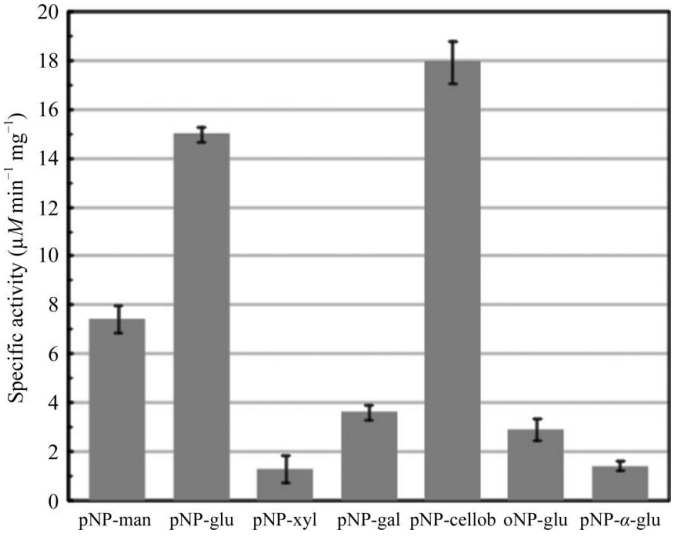

Figure 1
CsMan5 β-mannanase specific activity towards *p*-nitrophenyl substrates. Abbreviations for substrates are as follows: pNP-man, *p*-nitrophenyl β-mannopyranoside; pNP-glu, *p*-nitrophenyl β-glucopyranoside; pNP-xyl, *p*-nitrophenyl β-xylanopyranoside; pNP-gal, *p*-nitrophenyl β-galactopyranoside; pNP-cellob, *p*-nitrophenyl β-cellobioside; oNP-glu, *o*-nitrophenyl β-glucopyranoside; pNP-α-glu, *p*-nitrophenyl α-glucopyranoside.

Studies of the secretome of *N. crassa* have acknowledged this gene product as a secreted protein; however, an N-terminal signal peptide is not recognized in the sequence (*SignalP* 4.0 server; Sun & Glass, 2011; Tian *et al.*, 2009). Furthermore, the Fungal Secretome Knowledge Base (FunSecKB; Lum & Min, 2011), which uses a diverse computational protocol based on a variety of prediction algorithms, does not identify NCU08412 as a secreted protein. There are other cases similar to this in the *N. crassa* genome (NCU00972 and NCU01900) and similar situations have been reported for *Aspergillus* species (Medina *et al.*, 2005). It has been suggested that these proteins are secreted *via* a nonconventional/nonclassical secretory pathway (Medina *et al.*, 2005), although other authors have claimed that unrecognized signal peptides may arise from lesser known sequences and/or incorrect annotation of the start codon (Braaksma *et al.*, 2010). In fact, the *SecretomeP* server (http://www.cbs.dtu.dk/services/SecretomeP/) calculates an NN-score of 0.544 for NCU08412, which is barely above the threshold of 0.5 that identifies nonclassically secreted proteins. This hampered determination of the putative full sequence of the *C. sitophila* Man5 gene. Degenerate primers for gene amplification from a cDNA library generated from the mRNA purification were therefore designed based on homology sequence analysis of *N. crassa* endo-β-D-1,4-mannanase (NCU08412). The clarity of the electron-density maps obtained from the X-ray diffraction of CsMan5 crystals allowed safe assignment of the side chains of some residues encompassed within the designed primer sequence and thus assists in decreasing the degeneracy. Pure CsMan5 protein and protein crystals identical to those from which X-ray data were collected were submitted to N-terminal sequencing. The first nine residues sequenced (KVPKGFVTT) indicate that the protein isolated from the fermentation supernatant has been processed in the manner expected for a mature secreted protein.

Alignment of CsMan5 with the *N. crassa* homologue (not including the 14 residues at the N-terminus) and with the *H. jecorina* homologue (corresponding to the recombinant catalytic domain only) rendered 91.5 and 36.2% identity and 95.4 and 50.3% similarity, respectively (Fig. 2). Sequencing of the amplified gene isolated from a generated cDNA library showed that CsMan5 has three insertions (Glu196, Glu236 and Asp237) compared with the *N. crassa* homologue.

3.2. Overall structure description and comparison with other GH5 and GH26 members

The crystal structure of CsMan5 was determined by molecular replacement at 1.40 Å resolution. The final model,

consisting of 381 residues, has a crystallographic R factor of 16.1% and a R_{free} value of 20.2%. The overall structure of CsMan5 is represented in Figs. 3(*a*) and 3(*b*). The protein molecule, with approximate dimensions of 56 × 50 × 48 Å, exhibits a single-domain architecture composed exclusively of a catalytic domain. The structure reveals the classical $(\beta/\alpha)_8$-barrel fold typical of the family 5 glycosyl hydrolases. It also features a roughly V-shaped cleft that by structural homology with other GH5 members corresponds to the catalytic groove in which the catalytic acid/base and nucleophile (Glu181 and Glu301) are located. This cleft is sealed by two short β-strands (β1 and β2) from the N-terminus, a feature that is observed in the structures of related GH5-family enzymes (Fig. 3*a*).

The *C. sitophila* enzyme shows two different *N*-glycosylation sites. It was possible to model one site bound to Asn63 which contains two linked *N*-acetylglucosamine residues. This glycosylation may have a structural role in stabilizing structural elements, as there are hydrogen bonds between the glucosamine residues and Ile67 (backbone N—H) and Leu71 (backbone C═O), which both belong to a very long loop (residues 64–84), and Val136 (C═O) located on helix 4. The other site (Asn319) is not fully occupied, suggesting disorder or perhaps different populations of glycosylation. Neither of these sites coincide with the four sites reported for the *H. jecorina* enzyme (HjMan5A; Sabini *et al.*, 2000). Both *A. nidulans* Man5C (Dilokpimol *et al.*, 2011) and the *N. crassa* homologue have two predicted *N*-glycosylated sites, one of which corresponds to Asn63 of CsMan5 in terms of sequence homology.

The first four residues confirmed by sequencing were not visible in the electron density; residues Glu236 and Asp237 were also not well defined, suggesting flexibility of the region, and were therefore not modelled. Residue Ala186 has 'disallowed' Ramachandran angles. This amino acid belongs to an extended loop that is absent in HjMan5A; however, the electron density is very well defined, leaving no doubt regarding the interpretation of this conformation. The structure also contains a chloride ion, two acetate molecules and one Tris molecule retained from the crystallization mother liquor; the latter is positioned in the −1 site, which is a common occurrence in glycoside hydrolases (James & Lee, 1996; Sabini *et al.*, 2000).

The overall fold of CsMan5 is strongly conserved when compared with other endo-β-D-1,4-mannanase structures described to date and has highest similarity to those within the GH5 family. Superimposition with the HjMan5A structure (PDB entry 1qno), the only structure of fungal origin

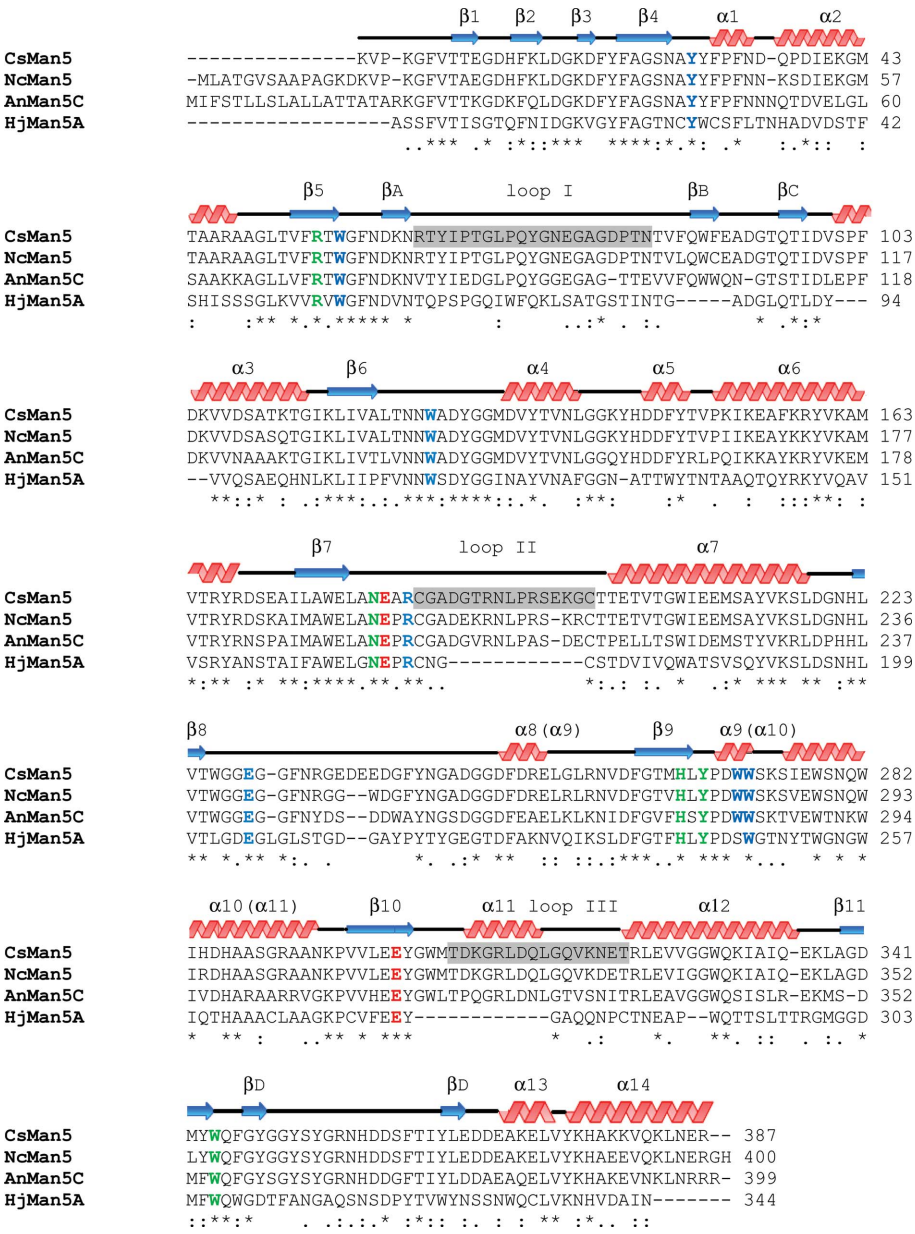

Figure 2
Sequence alignment of *C. sitophila* endo-β-D-1,4-mannanase with the homologues from *N. crassa* (92% identity), *A. nidulans* FGSC A4 (71% identity) and *H. jecorina* (36% identity). Helices and β-sheets are numbered according to the HjMan5A structure; where numbering diverges the HjMan5A nomenclature is given in parentheses. The strictly conserved catalytic residues are highlighted in red; the five residues conserved among mannanases and cellulases belonging to the GH5 family are highlighted in green. Residues participating in the substrate orientation and hydrophobic/stacking platform in the binding cleft are shown in blue. Residues from CsMan5 belonging to the extended-loop regions are highlighted with a grey background; loop III includes helix 11. Asterisks (*) denote identical residues, colons (:) denote very similar residues and periods (.) denote similar residues. The multiple sequence alignment was performed with *ClustalW* v.2.1 (Larkin *et al.*, 2007).

determined to date and the most similar according to the *DALI* and *HHpred* servers (Holm & Rosenström, 2010; Söding *et al.*, 2005), gave a root-mean-square deviation (r.m.s.d.) of 1.46 Å for 381 equivalent C^{α} atoms using *LSQMAN* (Kleywegt & Jones, 1994). Along with 13 well conserved residues within the vicinity of the active site, there is a nonprolyl *cis*-peptide bond between residues Trp344 and Gln345 equivalent to those observed in the active sites of *Thermonospora fusca*, *Mytilus edulis* (blue mussel) and *H. jecorina* β-mannanases. This feature is essential for the correct conformation of Trp344, an important residue in the active-site architecture (Sabini *et al.*, 2000; Hilge *et al.*, 1998; Larsson *et al.*, 2006). In spite of these similarities, there are still some noteworthy differences. CsMan5 has only one disulfide bond, between Cys184 and Cys199 (at the end of strand 7 and at the beginning of helix α7, respectively), which superimposes well with one of the three disulfide bonds in the HjMan5A structure. The cystine bond in CsMan5 is positioned at the base of an extensive loop of 14 residues; this structural element is not represented in HjMan5A. This loop (loop II; Cys184–Cys199) is in the vicinity of the cleft accessing the active site (Fig. 3*b*). The two other major structural differences also concern loops flanking the catalytic groove. They are located between residues Arg64 and Asn84 (loop I) and between residues Thr306 and Thr321 (loop III). Loop I has a prolyl *cis*-peptide bond between Leu71 and Pro72 that introduces a kink into this structural element, projecting the loop over the entry to the cleft. Loop III comprises a short α-helix between residues Asp307 and Gly315. Together, these three loops seem to increase the depth of the catalytic cleft compared with that observed in *H. jecorina* (PDB entry 1qno) and render the entry narrower (Fig. 4*a*). These three elements are also not present in the crystal structure of tomato endo-β-D-1,4-mannanase (*Lycopersicon esculentum* or *Solanum lycopersicum*; PDB entry 1rh9; Bourgault *et al.*, 2005), which has a sequence identity of 27% and an r.m.s.d. of 1.61 Å (for all C^{α} pairs) when superimposed with CsMan5, or in the structure of blue mussel endo-β-D-1,4-mannanase (*M. edulis*; PDB entry 2c0h; Larsson *et al.*, 2006), which has 18% sequence identity and an r.m.s.d. of 2.70 Å for all C^{α} pairs. Both of these proteins represent true endo enzymes with an open groove across the enzyme surface.

Among the determined structures of GH5 enzymes of prokaryotic origin, few show a sequence identity with CsMan5 of greater than 20%. The most similar are the endo-β-D-1,4-mannanase from *Thermotoga petrophila* (34% sequence identity; r.m.s.d. = 1.21 Å; PDB entry 3pzg; Ramos dos Santos *et al.*, 2012) and the exo-β-mannosidase from *Cellvibrio mixtus* (22% sequence identity; r.m.s.d. = 1.40 Å; PDB entry 1uuq; Dias *et al.*, 2004). Among the prokaryotic β-mannanase structures belonging to the GH5 family that are available [PDB entries 1bqc, 2man and 3man from *T. fusca* (16% sequence identity; Hilge *et al.*, 1998), 1wky from *Bacillus* sp. strain JAMB-602 (17% sequence identity; Akita *et al.*, 2004), 3civ from *Alicyclobacillus acidocaldarius* (20% sequence identity; Zhang *et al.*, 2008), 2whj from *B. agaradhaerens* (16% sequence identity; Tailford *et al.*, 2009), 3jug from *Bacillus* sp. N16-5 (18% sequence identity; Zhao *et al.*, 2011) and 3pzg from *T. petrophila* (Ramos dos Santos *et al.*, 2012)], none exhibit the extended loops at the top of the β-barrel equivalent to those described for CsMan5 (Fig. 4*b*). Only the exo-cleaving enzyme *C. mixtus* mannosidase 5A (CmMan5A) presents equivalent loops in terms of structure, although only one of three coincides in terms of sequence (Dias *et al.*, 2004). The *C. mixtus* mannanase 26C (CmMan26C; 11% sequence identity; r.m.s.d. = 3.33 Å; PDB entry 2vx4) likewise exhibits exo activity, preferentially releasing mannobiose from the nonreducing end of mannan and mannooligosaccharides (Cartmell *et al.*, 2008). This enzyme also displays extended loops that affect the arrangement of the substrate-binding cleft. In particular, an extension of a surface loop creates a

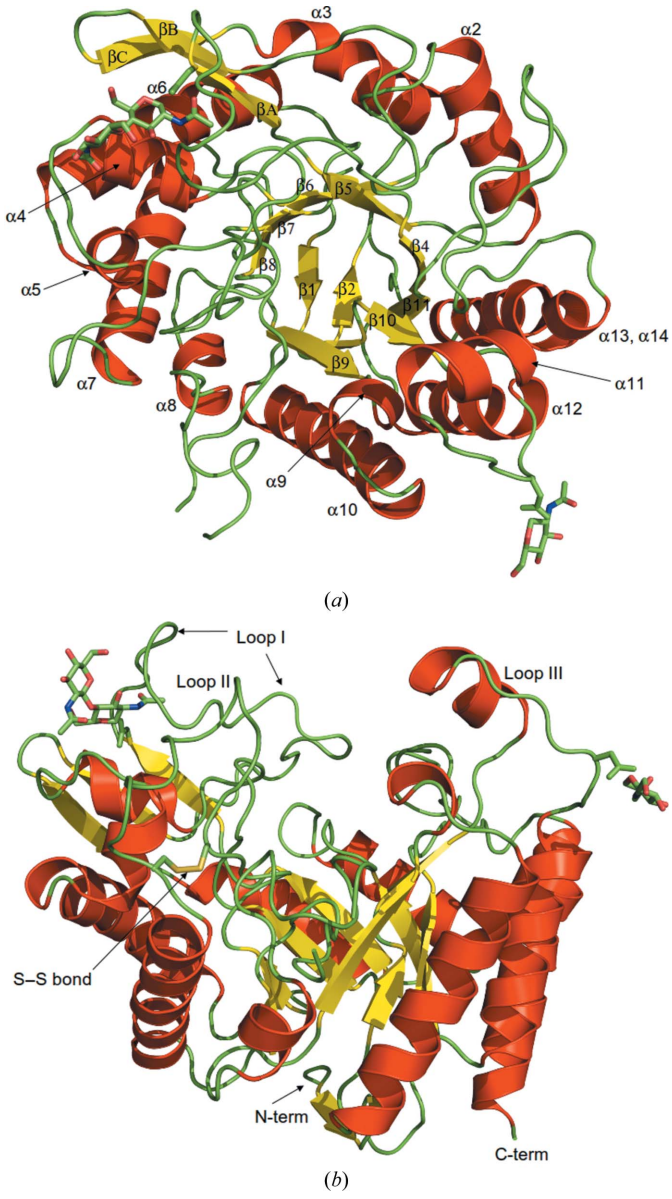

Figure 3
(*a*) Overall tertiary structure of *C. sitophila* β-mannanase showing the $(\beta/\alpha)_8$ fold from a top-down view. The two glycosylation sites are shown. (*b*) Side view of the structure of *C. sitophila* β-mannanase showing the extended loop regions (I, II and III).

steric block that limits the size of the cleft for substrate accommodation to four mannose-binding subsites.

As suggested previously, the extended loops on CmMan5A modified the active-centre accessibility and modulated the specificity from endo to exo (Dias *et al.*, 2004). The extended loops found in the CsMan5 structure seem to contribute to the enclosure of the cleft (loop I, residues 64–84), overlapping with two of the extended sequences (residues 88–97 and 378–412) in CmMan5A. This structural element provides steric blockage on one side of the cleft that is entirely absent in HjMan5 (Fig. 4) but is less dramatic than that in CmMan5A. The loop between $\beta7$ and $\alpha7$ (loop II, residues 184–199) provides additional steric blockage and is somewhat spatially equivalent to the loop composed of residues 131–171 of CmMan5A identified as providing the 'pocket' nature of the site and conferring the true exo activity. The third extended loop of CsMan5 (loop III, residues 306–321), which is positioned opposite to loop II, may also offer further steric hindrance to large branched polysaccharides, as it extends over the cleft, making it deeper and narrower (Fig. 4). Finally, it should be noted that the loop connecting $\beta5$ and $\alpha5$ at the extreme reducing end of the aglycone subsites is spatially equivalent to that in the HjMan5A structure, a common feature observed in other endo-mannanases.

3.3. Comparison of the active site

Pairwise structural alignments have shown that the structure of CsMan5 belongs to family 5 of the GH-A clan. Enzymes in this clan possess an active-site pocket with highly conserved features (Durand *et al.*, 1997). CsMan5 complies with these features, revealing two pivotal catalytic glutamate residues (Glu181, the acid/base, and Glu301, the nucleophile) positioned in equivalent positions as in many GH5 members. CsMan5 was isolated from a fungal culture grown solely on microcrystalline cellulose, which was expected to induce endoglucanase expression. Of the chromogenic substrates shown in Fig. 1, the highest activities were towards *p*-nitrophenyl β-cellobioside and *p*-nitrophenyl β-glucopyranoside. However, CsMan5 also showed enzymatic activity towards a derivatized mannan. It is well established through structural analyses of GH5 cellulases and GH mannanases that similar active sites are used for the hydrolysis of both substrates (Hilge *et al.*, 1998; Sakon *et al.*, 1996). Nevertheless, several strictly conserved residues identified as key residues within the active site of β-mannanases are present in CsMan5 and superimpose with the same amino acids in HjMan5A: Arg55 (Arg54), Trp125 (Trp114), Asn180 (Asn168), His266 (His241) and Tyr268 (Tyr243) (HjMan5A residues are shown in parentheses). The catalytic proton donor Glu181 is involved in hydrogen bonds to His266 and Trp125 (Fig. 5*a*). It has previously been suggested and shown that the imidazole N atom of the equivalent of His266 contributes to the positioning and ionization of the catalytic glutamate (Sabini *et al.*, 2000; Larsson *et al.*, 2006). The orientation of the nucleophile glutamate Glu301 is provided through a hydrogen bond to Tyr268, which is highly conserved except in CmMan5A, which contains a tryptophan at this position; the other O atom of the carboxylate Glu301 is within hydrogen-bonding distance of Asn180 and Arg55 (Fig. 5*a*). Arg55 is highly conserved in many reported structures of β-mannanases (PDB entries 1bqc, 1rh9, 2c0h, 1uuq, 1qno and 3pzg) and in the GH-A clan in general. It plays an important role in conferring structural stability to the active site by forming a salt bridge with Glu301 and a hydrogen bond to Asn180. In the present structure, a very well defined Tris molecule is bound in the −1 site, forming hydrogen bonds to the two catalytic glutamate residues, Asn180, three water molecules and an acetate molecule

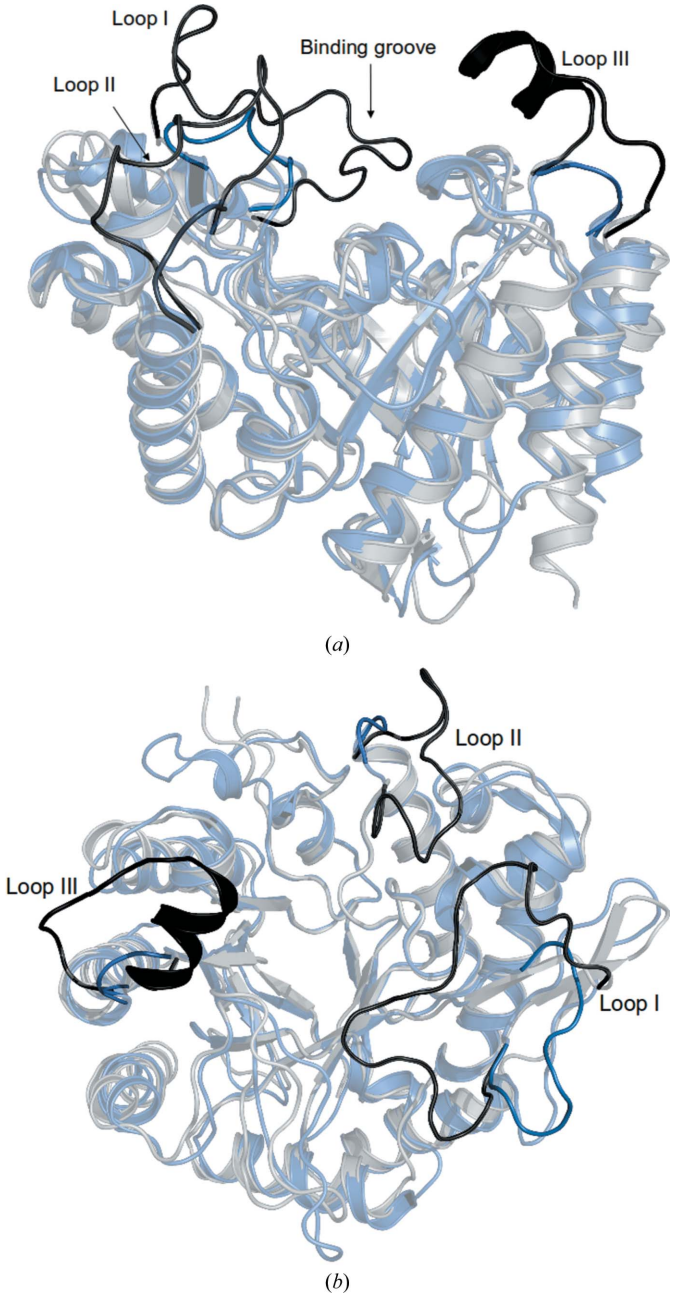

Figure 4
(*a*) Side view and (*b*) top view of the superimposition of *H. jecorina* Man5A (blue) on CsMan5 (grey). The extended loops in CsMan5 are shaded black, while the equivalent loops in HjMan5A are shaded dark blue.

(Fig. 5). This arrangement, apart from the acetate, was also observed for HjMan5A (PDB entry 1qno). The distance between Glu181 and Glu301 is 3.90 Å, which is a clear indication that the hydrolytic catalysis is based on the double-displacement retention mechanism (Durand *et al.*, 1997).

3.4. Substrate-binding residues

The binding-site geometry in CsMan5 maintains all of the absolutely conserved residues that are essential to substrate recognition and orientation for catalysis in GH5 members. These residues, Arg55, Asn180, Glu181, His266, Tyr268, Glu301 and Trp344, encompass subsite −1 (Fig. 5*a*). Similarities between the active sites of CsMan5 and other mannanases solved in complex with two or three units of mannan polysaccharide (*i.e.* mannobiose or mannotriose) allowed the mapping and assignment of other substrate-binding sites with confidence. The −2 subsite is surrounded by the conserved amino acids Tyr29, Trp57 (which may also offer polar contacts at the −1 subsite) and Asp359 (Fig. 5*b*), identical to what is observed in HjMan5A (Tyr27, Trp56 and Asp321; the latter is structurally conserved as Asn259 in TfMan5A) and also in the β-mannanase from *T. petrophila* (Tyr45, Trp73 and Asp371; Ramos dos Santos *et al.*, 2012). On the other side of the active site, Asp127 may contribute to further stabilize the sugar through hydrogen bonding (Figs. 5*a*

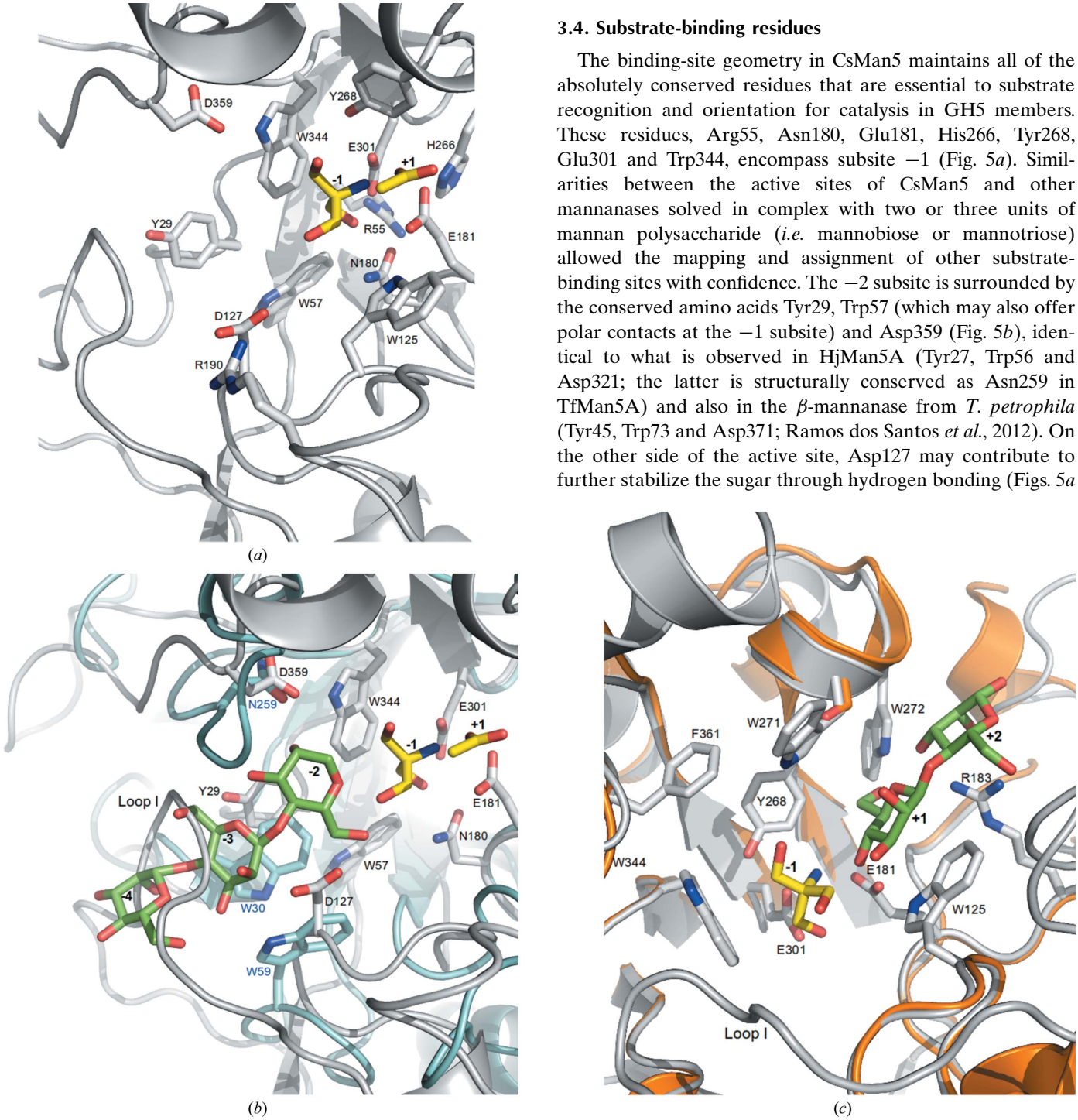

Figure 5
Representation of the active site and substrate-binding cleft, highlighting the glycone and aglycone subsites. (*a*) The CsMan5 structure (grey) complexed with Tris and acetate (ACT) (subsites −1 and +1, respectively; yellow), depicting the major highly conserved residues in the GH5 family: Arg55, Trp125, Asn180, Glu181, His266, Tyr268, Glu301 and Trp344. (*b*) The CsMan5 structure (grey) complexed with Tris and ACT is superimposed with *T. fusca* Man5 (cyan) in complex with mannotriose in subsites −2, −3 and −4. The mannotriose ligand (subsites −2, −3 and −4) is shown in green, while the Tris and ACT molecules (subsites −1 and +1) are shown in yellow. The positioning of loop I shows an evident clash with subsites −3 and −4. (*c*) Superimposition of CsMan5 (grey) complexed with Tris in the −1 subsite with *H. jecorina* Man5A (orange) in complex with mannobiose (green) at subsites +1 and +2. The acetate molecule (ACT) represented in (*b*) would superimpose with the mannose in subsite +1. The catalytic glutamates Glu181 and Glu301 are coordinated to the Tris at the −1 subsite, together with the strictly conserved Trp344. Ser246 (orange) from HjMan5A is shown to superimpose well with Trp271.

and 5*b*). This residue is oriented by Arg190, which belongs to loop II. Although CsMan5 has the same conserved Tyr128 as HjMan5A (Tyr117) that may compose the −3 subsite (equivalent to Tyr30 in TfMan5A), it can be easily observed that this site is entirely blocked by the backbone of loop I (in particular residues Asn76–Asp81). The positioning of this loop does not seem to be a packing effect or a crystallization artefact. In fact, this extended element makes many hydrogen-bond contacts with different structural elements of CsMan5 (α1, α3 and the loop between β6 and α4) and contains two linked *N*-acetylglucosamine residues that provide rigidity to the loop (the average *B* factor of the main chain of the 21 residues in this loop is approximately 21 Å^2). This feature may be relevant in determining substrate specificity. In this respect, loop I from CsMan5 has a similar effect as the surface loop from CmMan26C (Cartmell *et al.*, 2008) that creates a steric block at the distal glycone −2 subsite.

The hyperthermophilic endoglucanase Cel5A from *T. maritima* (17% sequence identity to CsMan5 and 2.35 Å r.m.s.d. over all C$^\alpha$ pairs; PDB entry 3mmw) has been reported to exhibit dual activity towards both glucan-based and mannan-based polysaccharides (Chhabra *et al.*, 2002; Wu *et al.*, 2011). Structural studies of TmCel5A (Wu *et al.*, 2011; Pereira *et al.*, 2010) allowed the identification of the strictly conserved residues in GH5 that are involved in substrate binding and catalysis around the −1 subsite. However, the −2 subsite shows relevant differences regarding the three conserved residues observed in endo-β-D-1,4-mannanases referred to above. Trp57 of CsMan5 aligns with His95 from TmCel5A and may establish polar interactions with a sugar unit, but there is no corresponding residue to His96 of TmCel5A in CsMan5, which is a residue that is involved in polar contacts with the −2 subsite. Residues Tyr29 and Asp359 have no sequence or structural homology in TmCel5A. Superimposition of CsMan5 with TmCel5A in complex with mannotriose (PDB entry 3azs; T.-H. Wu, C.-H. Huang, T.-P. Ko, H.-L. Lai, Y. Ma, C.-C. Chen, Y.-S. Cheng, J.-R. Liu & R.-T. Guo, unpublished work) shows that Asp359 is able to establish polar contacts with the mannose unit but would possibly clash with the equatorial C2 hydroxyl group of cellotetraose, suggesting a possible mechanism of selectivity. Also noteworthy is the fact that Trp210 in TmCel5A, which is important for providing stacking forces with the −2 subsite, is absent in CsMan5 (and also in HjMan5A as well as TpMan). The active-site architecture surrounding the −2 subsite in CsMan5 has more in common with other reported endo-β-D-1,4-mannanases than with this endo-glucanase with dual activity.

In the CsMan5 structure the +1 subsite is occupied by an acetate molecule (Fig. 5*b*). This molecule establishes bonds to some of the relevant side chains identified in substrate recognition in HjMan5A (PDB entry 1qnr). Trp125 (Trp114 in HjMan5A) is a conserved residue that by comparison should be involved in hydrophobic stacking, while Trp272 (Trp247) forms a hydrogen bond to the (OH)6 H atom of the mannosyl ring (in the present structure it binds to the acetate molecule; Fig. 5*c*). Further similarities can be tracked regarding substrate-binding residues in the +2 subsite: superimposition of CsMan5 with HjMan5A shows the well conserved Arg183 and Trp272 that are important in positioning the +2 ring (Fig. 5*c*). Another essential residue is Glu229. It is remarkable that these three residues superpose so well with their equivalents in HjMan5A, particularly when the +2 subsite is entirely vacant in the present structure. This implies that the described residues offer a well structured and rather stiff scaffold for substrate recognition. Substantial differences surrounding the +1 and +2 subsites can be found in *T. maritima* endo-glucanase: there is no tryptophan corresponding to Trp272 (there is a neighbouring phenylalanine, Phe201) and the hydrophobic platform for the +1 subsite corresponding to Trp125 is spatially displaced (Trp173 in TmCel5A).

There is one noteworthy difference in this region of the binding cleft of CsMan5: aligned with the +1 subsite and opposite to Trp125 is another tryptophan residue, Trp271, which apparently offers a hydrophobic platform for substrate interactions (Fig. 5*c*). This residue is perfectly aligned with Ser246 from HjMan5A and is otherwise not conserved in any of the β-mannanases reported to date. The *T. maritima* Cel5A and *T. petrophila* Man5A structures exhibit a nonconserved histidine in this position. In the latter this residue adopts a well defined rotamer conformation when complexed with a sugar, suggesting a role in substrate recognition and interaction. The neighbouring residues Tyr268 (strictly conserved) and Phe361 (Tyr323 in HjMan5A) compose the hydrophobic moiety. Although Trp271 does not establish any bonds with the acetate molecule in the aglycone +1 subsite, it is plausible that contacts may arise with a bulkier molecule such as a mannose sugar ring which are impossible with the serine residue in HjMan5A.

Recently, two endo-β-1,4-mannanases from *A. nidulans* FGSC A4, Man5A and Man5C, have been functionally characterized (Dilokpimol *et al.*, 2011). CsMan5 has 36% identity to AnMan5A, while its identity to AnMan5C is strikingly high at 71%. The high identity between CsMan5 and AnMan5C includes the sequences encompassed by the extended loops (Fig. 2). These two endo-mannanases were enzymatically characterized and both showed increased catalytic efficiency towards manno-oligosaccharides with increasing degrees of polymerization (DP; mannotetraose, M4, to mannohexaose, M6). However, important differences in specificity and kinetic affinities were noted: AnMan5A showed a specific activity towards locust bean gum galactomannan similar to the reported activities of GH5 endo-mannanases such as HjMan5A. AnMan5C showed a higher specific activity (30–80% higher) towards glucomannan and galactomannans than AnMan5A and, most interestingly, prefers polysaccharides with fewer α-1,6-galactosyl substituents, as opposed to AnMan5A which appeared to be insensitive to changes in the backbone composition and to substitution in the mannan substrate. The study described AnMan5A as having a more kinetically relevant −3 subsite than AnMan5C. The latter showed a difference in hydrolysis of M4 substrates, with M2 being the major hydrolysis product, thus suggesting important differences in the glycone-binding region. This finding is in accordance with the structural evidence that we have

encountered in the crystal structure of CsMan5. Although 4-nitrophenylglycosides are not exclusively cleaved by exo-acting enzymes and are not the most informative substrates for elucidating specificity, the preference for 4-nitrophenyl-β-D-cellobioside observed in preliminary enzymatic assays (Fig. 1) may corroborate this feature. Further and more specific assays are needed to clarify the role of the -2 subsite. Furthermore, Dilokpimol *et al.* (2011) went on to prove using site-directed mutagenesis that the tryptophan (Trp283 in AnMan5C) at the +1 subsite, equivalent to Trp271 in CsMan5, played a crucial role in the enhanced transglycosylation activity; they also showed that transglycosylation yields could be increased by 50% by replacing tryptophan by a serine at the +1 subsite in AnMan5A. We propose that Trp271 in CsMan5 may play an important role in discrimination between mannan substrates with different extents of branching. Inspection of the structure of TmMan5 suggests that a α-1,6-galactosyl moiety could not be accommodated in the +1 subsite, whereas acetyl substitutions at O2 and O3 are sterically possible. The average distance between the two tryptophan residues that provide the hydrophobic stacking of the +1 subsite (Trp125 and Trp271) is 9.5 Å, making bulkier branched substrates difficult to accommodate. Trp272 is solvent-accessible and is flanked by the extended loop III; on the opposite side of the cleft, loop II flanks the opening. We postulate that these features contribute to a dramatic change in the active-site accessibility and possibly contribute to modulating the enzyme specificity.

4. Conclusion

The determination of the three-dimensional structure of *C. sitophila* endo-β-D-1,4-mannanase revealed unexpected details of what would have been an unremarkable conserved $(\beta/\alpha)_8$-barrel fold of the GH5 family based solely on model predictions. The results presented here suggest that CsMan5 has mixed activity towards glucan-based and mannan-based polysaccharides. However, CsMan5 shows the highest sequence and structural homology to the β-mannanases. More importantly, the architecture conferred by the residues surrounding the -2 and +1 subsites in CsMan5 is strikingly similar to those reported for the *H. jecorina* Man5A and *T. petrophila* Man5 structures, whereas the endo-glycanase *T. maritima* Cel5A exhibits divergence regarding these subsites. We were able to identify four subsites in CsMan5: -2, -1, +1 and +2. There is also the possibility that CsMan5 may present some exo-mannanase activity, similar to that reported for CmMan5A from *C. mixtus*, and that, like CsMan26C, CsMan5 is a 2-mannobiohydrolase considering the limited four mannose-binding subsites. Further enzymatic studies aiming at a thorough substrate-specificity characterization, in particular of transglycosylation activity, are ongoing. Further enzymatic and structural studies involving mutational manipulations will allow the determination of the role of the extended loops flanking the binding cleft in discriminating (branched) substrates and the possible presence of a +3 subsite as well as the role of Trp272. CsMan5 is an interesting example of evolutionary specialization circumscribed to the loops surrounding the active site of a common and widespread fold in nature. This endo-mannanase with mixed activity towards glucan-based polysaccharides and galacto-derivatized mannans (and perhaps other polysaccharides) is the first reported structure of this family of eukaryotic origin to show a variation in the substrate-binding cleft with an impact on substrate accessibility and possibly specificity.

Maria Arménia Carrondo is gratefully acknowledged for support. The European Synchrotron Radiation Facility in Grenoble, France is kindly acknowledged for the provision of synchrotron-radiation facilities and support. Sandra Viegas from ITQB is thanked for support in mRNA manipulation. Sonia Vitorino, Marisa Caeiro and Cristina Silva Pereira are gratefully acknowledged for helpful discussions. CSS holds a PhD fellowship from FCT, Portugal (SFRH/BD/40586/2007). This work was supported by Fundação para a Ciência e a Tecnologia through grant PEst-OE/EQB/LA0004/2011.

References

Akita, M., Takeda, N., Hirasawa, K., Sakai, H., Kawamoto, M., Yamamoto, M., Grant, W. D., Hatada, Y., Ito, S. & Horikoshi, K. (2004). *Acta Cryst.* D**60**, 1490–1492.
Arx, J. A. von (1981). *Sydowia*, **34**, 13–29.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Bourgault, R., Oakley, A. J., Bewley, J. D. & Wilce, M. C. (2005). *Protein Sci.* **14**, 1233–1241.
Braaksma, M., Martens-Uzunova, E. S., Punt, P. J. & Schaap, P. J. (2010). *BMC Genomics*, **11**, 584.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Cartmell, A., Topakas, E., Ducros, V. M.-A., Suits, M. D. L., Davies, G. J. & Gilbert, H. J. (2008). *J. Biol. Chem.* **283**, 34403–34413.
Centeno, S. & Calvo, M. A. (2001). *Microbios*, **106**, 69–73.
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.
Chhabra, S. R., Shockley, K. R., Ward, D. E. & Kelly, R. M. (2002). *Appl. Environ. Microbiol.* **68**, 545–554.
DeBoy, R. T., Mongodin, E. F., Fouts, D. E., Tailford, L. E., Khouri, H., Emerson, J. B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H. J. & Nelson, K. E. (2008). *J. Bacteriol.* **190**, 5455–5463.
Dias, F. M., Vincent, F., Pell, G., Prates, J. A., Centeno, M. S., Tailford, L. E., Ferreira, L. M., Fontes, C. M., Davies, G. J. & Gilbert, H. J. (2004). *J. Biol. Chem.* **279**, 25517–25526.
Dilokpimol, A., Nakai, H., Gotfredsen, C. H., Baumann, M. J., Nakai, N., Abou Hachem, M. & Svensson, B. (2011). *Biochim. Biophys. Acta*, **1814**, 1720–1729.
Durand, P., Lehn, P., Callebaut, I., Fabrega, S., Henrissat, B. & Mornon, J.-P. (1997). *Glycobiology*, **7**, 277–284.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Gilbert, H. J., Stålbrand, H. & Brumer, H. (2008). *Curr. Opin. Plant Biol.* **11**, 338–348.
Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P. & Davies, G. J. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 7090–7094.
Hilge, M., Gloor, S. M., Rypniewski, W., Sauer, O., Heightman, T. D., Zimmermann, W., Winterhalter, K. & Piontek, K. (1998). *Structure*, **6**, 1433–1444.
Hogg, D., Pell, G., Dupree, P., Goubet, F., Martín-Orúe, S. M., Armand, S. & Gilbert, H. J. (2003). *Biochem. J.* **371**, 1027–1043.

Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
James, J. A. & Lee, B. H. (1996). *Biotechnol. Lett.* **18**, 1401–1406.
Kleywegt, G. J. & Jones, T. A. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 9–14.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2980.
Larsson, A. M., Anderson, L., Xu, B., Muñoz, I. G., Usón, I., Janson, J. C., Stålbrand, H. & Ståhlberg, J. (2006). *J. Mol. Biol.* **357**, 1500–1510.
Lum, G. & Min, X. J. (2011). *Database*, **2011**, bar001.
Martinez, D. *et al.* (2008). *Nature Biotechnol.* **26**, 553–560.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Medina, M. L., Haynes, P. A., Breci, L. & Francisco, W. A. (2005). *Proteomics*, **5**, 3153–3161.
Moreira, L. R. & Filho, E. X. (2008). *Appl. Microbiol. Biotechnol.* **79**, 165–178.
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.
Oliveira, A. C., Peres, C. M., Correia Pires, J. M., Silva Pereira, C., Vitorino, S., Figueiredo Marques, J. J., Barreto Crespo, M. T. & San Romão, M. V. (2003). *Microbiol. Res.* **158**, 117–124.
Park, S. H., Park, K. H., Oh, B. C., Alli, I. & Lee, B. H. (2011). *N. Biotechnol.* **28**, 639–648.
Pereira, J. H., Chen, Z., McAndrew, R. P., Sapra, R., Chhabra, S. R., Sale, K. L., Simmons, B. A. & Adams, P. D. (2010). *J. Struct. Biol.* **172**, 372–379.
Ramos dos Santos, C., Paiva, J. H., Meza, A. N., Cota, J., Alvarez, T. M., Ruller, R., Prade, R. A., Squina, F. M. & Murakami, M. T. (2012). *J. Struct. Biol.* **177**, 469–476.
Rosa, M. E., Pereira, H. & Fortes, M. (1990). *Wood Fiber Sci.* **22**, 149–169.
Sabini, E., Schubert, H., Murshudov, G., Wilson, K. S., Siika-Aho, M. & Penttilä, M. (2000). *Acta Cryst.* D**56**, 3–13.
Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R. & Karplus, P. A. (1996). *Biochemistry*, **35**, 10648–10660.
Sanctis, D. de *et al.* (2012). *J. Synchrotron Rad.* **19**, 455–461.
Shear, C. L. & Dodge, B. O. (1927). *J. Agric. Res.* **34**, 1019–1042.
Silva Pereira, C., Pires, A., Valle, M. J., Vilas-Boas, L., Figueiredo Marques, J. J. & San Romão, M. V. (2000). *J. Ind. Microbiol. Biotechnol.* **24**, 256–261.
Silva Pereira, C., Soares, G. A. M., Oliveira, A. C., Rosa, M. E., Pereira, H., Moreno, N. & San Romão, M. V. (2006). *Int. Biodeterior. Biodegradation*, **57**, 244–250.
Söding, J., Biegert, A. & Lupas, A. N. (2005). *Nucleic Acids Res.* **33**, W244–W248.
Sternberg, D. (1976). *Biotechnol. Bioeng. Symp.*, pp. 35–53.
Sun, J. & Glass, N. L. (2011). *PLoS One*, **6**, e25654.
Tailford, L. E., Ducros, V. M.-A., Flint, J. E., Roberts, S. M., Morland, C., Zechel, D. L., Smith, N., Bjørnvad, M. E., Borchert, T. V., Wilson, K. S., Davies, G. J. & Gilbert, H. J. (2009). *Biochemistry*, **48**, 7009–7018.
Tian, C., Beeson, W. T., Iavorone, A. T., Sun, J., Marletta, M., Cate, J. H. & Glass, N. L. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 22157–22162.
Vitorino, S. I., Neves, E. S., Gaspar, F., Figueiredo Marques, J. J. & San Romão, M. V. (2007). *Ciência Téc. Vitiv.* **1**, 1–4.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wu, T.-H., Huang, C.-H., Ko, T.-P., Lai, H.-L., Ma, Y., Chen, C.-C., Cheng, Y.-S., Liu, J.-R. & Guo, R.-T. (2011). *Biochim. Biophys. Acta*, **1814**, 1832–1840.
Zhang, Y., Ju, J., Peng, H., Gao, F., Zhou, C., Zeng, Y., Xue, Y., Li, Y., Henrissat, B., Gao, G. F. & Ma, Y. (2008). *J. Biol. Chem.* **283**, 31551–31558.
Zhao, Y., Zhang, Y., Cao, Y., Qi, J., Mao, L., Xue, Y., Gao, F., Peng, H., Wang, X., Gao, G. F. & Ma, Y. (2011). *PLoS One*, **6**, e14608.