Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The three-dimensional structure of a *Trichoderma reesei* β-mannanase from glycoside hydrolase family 5

Elisabetta Sabini,[a] Heidi Schubert,[a] Garib Murshudov,[a] Keith S. Wilson,[a]* Matti Siika-Aho[b] and Merja Penttilä[b]

[a]Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, England, and [b]VTT Biotechnology and Food Research, PO Box 1501, FIN-02044 VTT Espoo, Finland

Correspondence e-mail: keith@yorvic.york.ac.uk

Received 9 September 1999
Accepted 29 October 1999

PDB References: β-mannanase–Pt-terpyridine, 1qnq; β-mannanase (room temperature), 1qno; β-mannanase (110 K, TRIS molecule in active site), 1qns; β-mannanase (110 K), 1qnp; β-mannanase–mannobiose, 1qnr.

The crystal structure of the catalytic core domain of β-mannanase from the fungus *Trichoderma reesei* has been determined at a resolution of 1.5 Å. The structure was solved using the anomalous scattering from a single non-isomorphous platinum complex with two heavy-metal sites in space group $P2_1$. The map computed with the experimental phases was enhanced by the application of an automated model building and refinement procedure using the amplitudes and experimental phases as observations. This approach is expected to be of more general application. The structure of the native enzyme and complexes with Tris–HCl and mannobiose are also reported: the mannobiose binds in subsites +1 and +2. The structure is briefly compared with that of the homologous β-mannanase from the bacterium *Thermomonospora fusca*.

1. Abbreviations

SAD, single anomalous diffraction; MIR, multiple isomorphous replacement; MAD, multiple-wavelength anomalous diffraction; SIRAS, single isomorphous replacement with anomalous scattering; MR, molecular replacement; r.m.s., root-mean-square; FOM, figure of merit; TRIS, Tris–HCl; NAG, *N*-acetyl glucosamine; kDa, kilodaltons; ADP, atomic displacement parameter; Pty, heavy-atom complex with Pt-terpyridine; Nat^{RT}, native structure at room temperature; Nat^{CC}_{TRIS}, structure of the native enzyme recorded under cryogenic conditions at 110 K with a TRIS molecule in the active site; Nat^{CC}, structure of TRIS-free native enzyme under cryogenic conditions at 110 K; Man2, enzyme complexed with mannobiose.

2. Introduction

2.1. Mannanases

Fungal plant polysaccharide-hydrolysing enzymes, cellulases and hemicellulases, are industrially important with a wide spectrum of well established applications. In recent years, hemicellulases have assumed growing interest owing to their potential application in the food/feed and pulp/paper industries (Viikari, Tenkanen *et al.*, 1993). Attention in the pulp and paper industry had been focused on xylanases because of the demonstration that loading of chlorine-containing chemicals during the bleaching of pulp can be reduced by addition of xylanases to the process (Viikari *et al.*, 1987). However, β-mannanases from the filamentous fungus *T. reesei* have also been shown to facilitate bleaching (Buchert *et al.*, 1993; Viikari, Kantelinen *et al.*, 1993).

Endo-1,4-β-D-mannanases (E.C. 3.2.1.78) catalyse random hydrolysis of β-D-1,4-mannopyranosyl linkages within the

main chain of mannans and various polysaccharides consisting mainly of mannose, such as glucomannans, galactomannans and galactoglucomannans. Usually a degree of polymerization (DP) of at least four is required, although some endo-mannanases are also able to hydrolyse mannotriose. The extent of hydrolysis of galactomannans by mannanases depends on the degree of substitution and the distribution of the substituents. Hydrolysis of glucomannans is affected by the proportion of glucose and mannose units as well as by the degree of acetylation (McCleary, 1991). β-Mannanases are produced by plants, bacteria and fungi. The endo-mannanase from *T. reesei* has been purified and characterized and mainly consists of several acidic pI forms (Stålbrand *et al.*, 1993) encoded by the *man1* gene (Stålbrand *et al.*, 1995).

Among the 77 families into which the catalytic domains of glycosyl hydrolases have been classified to date (http://afmb.cnrs-mrs.fr/~pedro/CAZY; Henrissat, 1991), the β-mannanase studied here belongs to family 5, clan GH-A. This family contains many enzymes involved in the microbial degradation of cellulose and hemicellulose, such as endo-glucanases (E.C. 3.2.1.4), exo-1,3-glycanases (E.C. 3.2.1.58) and β-mannanases (E.C. 3.2.1.78), whose sequences are only distantly related to the *T. reesei* mannanase. The sequence identity among the members of family 5 is low and an additional classification into eight subfamilies, A1–A8, has been introduced (Béguin, 1990; Lo Leggio *et al.*, 1997; Hilge *et al.*, 1998). The identity between members of the same subfamily is usually 25% or more and between subfamilies is rarely greater than 20%. Subfamily A7 contains eukaryotic mannanases, while subfamily A8 is composed of bacterial mannanases.

The *T. reesei* β-mannanase consists of two functional domains: a catalytic core joined to a putative substrate-binding domain by a linker unit (Stålbrand *et al.*, 1995). Such an organization is well established for a number of bacterial and fungal cellulases (Gilkes *et al.*, 1991). The *T. reesei* β-mannanase carries its polysaccharide-binding domain at the C-terminus, in contrast to the fungal cellulases belonging to the same hydrophobic cluster analysis family, which have N-terminal cellulase-binding domains (Schülein *et al.*, 1993). Attempts to crystallize the intact *T. reesei* mannanase proved unsuccessful, probably owing to the flexibility of the highly glycosylated linker; the structure of the fungal mannanase core is presented here. The core has an isoelectric point of 4.6 and a molecular weight of 40 kDa and was purified from a *T. reesei* that expresses a truncated mannanase containing the first 344 amino acids (Sabini *et al.*, 1999).

In family 5 of the *O*-glycosyl hydrolases, the structures of four cellulases had been published when we started this analysis: *Acidothermus cellulolyticus* endo-1,4-glucanase (Sakon *et al.*, 1996), *Bacillus agaradhaerans* endo-1,4-glucanase (Davies *et al.*, 1998), *Clostridium cellulolyticum* endo-1,4-glucanase A (Ducros *et al.*, 1995) and *C. thermocellum* endo-1,4-glucanase C (Dominguez *et al.*, 1996; Dominguez & Alzari, 1995). Given the low identity with the *T. reesei* β-mannanase, none of the then available coordinates proved to be a satisfactory model for molecular replacement (MR).

Subsequent to our elucidation of the *T. reesei* enzyme structure, the crystal structure of an additional family 5 member, that of the bacterial β-mannanase from *Thermomonospora fusca*, was published (Hilge *et al.*, 1998), together with its complexes with mannose oligomers. This showed that the structure had a $(\beta\alpha)_8$ barrel fold as expected from its homology to other family 5 members. The active site was shown to be a cleft running across the molecular surface and the positions of the two catalytic glutamates were identified, together with the sugar-binding subsites -2, -3 and -4. Here and in the rest of the paper, the sugar subsites are numbered according to a standard nomenclature (Davies *et al.*, 1997).

2.2. Experimental phasing

Experimental phases for macromolecular crystal structure determination have conventionally been generated by multiple isomorphous replacement (MIR) involving the binding of heavy atoms at a limited number of sites (Blow & Crick, 1959). The anomalous scattering components of the heavy atoms are routinely used to enhance the isomorphous information (MIRAS) or exploited in the combination of single isomorphous derivatives with their associated anomalous components (SIRAS; North, 1965). More recently, the determination of experimental phases has been extended to multiple-wavelength anomalous diffraction (MAD; Hendrickson & Ogata, 1997). The MIR technique continues to suffer from the non-isomorphism introduced into the crystal by the presence of the heavy atoms and different conditions of data collection such as precipitant conditions, pH or temperature. In this situation, the anomalous phasing component for an individual heavy-atom complex may be strong, but is not supported by any, or only by a weak, isomorphous signal. The usual approach has been to continue the search for derivatives which are more isomorphous with the native protein.

The present structure determination is of particular interest, as it depended solely on the anomalous scattering from a single non-isomorphous platinum complex of the β-mannanase core. The anomalous diffraction signal was sufficiently powerful to provide experimental phase information allowing the straightforward structure determination of the complex through automated building and refinement of a dummy-atom model, which provided phases close to those of the refined structure. The crystals were in the space group $P2_1$, with a single molecule in the asymmetric unit.

The native structure at 1.5 Å resolution was solved by placing the structure of the platinum derivative in the new cell, with the same space group and a small difference in unit-cell dimensions, and performing an initial rigid-body refinement. In addition, analysis of the structure of a complex with a mannan-derived oligosaccharide allows the identification of substrate-binding subsites +1 and +2. The general features of the structure are compared with those of the recently determined structure of the β-mannanase from *Therm. fusca* (Hilge *et al.*, 1998).

3. Experimental and computational methods

3.1. Crystallization

Crystals of *T. reesei* mannanase were grown at 293 K as described previously (Sabini *et al.*, 1999). In summary, an equal volume of 2 *M* ammonium sulfate buffered with 0.1 *M* Tris–HCl pH 8.5 was mixed with enzyme at 22 mg ml^{-1} stored in 10 m*M* TRIS pH 7.5. The resulting 4 µl drop was equilibrated over a well containing the ammonium sulfate/TRIS precipitant. The crystals belong to the monoclinic space group $P2_1$, with unit-cell dimensions $a = 50.01$, $b = 54.27$, $c = 60.24$ Å, $\beta = 111.31^\circ$. There is one molecule in the asymmetric unit, giving rise to a packing density of 1.90 Å^3 Da^{-1} corresponding to an approximate solvent content of 35% (Matthews, 1968).

The structure of the native protein proved to contain a molecule of TRIS bound in the active site (see below). Experiments to bind small oligomers of mannose to the TRIS-containing crystals were unsuccessful, the TRIS apparently preventing the binding of the substrate analogue. To obtain TRIS-free crystals, the protein was washed several times in 100 m*M* glycine pH 11 and stored in 20 m*M* glycine pH 8.0. Hanging drops were set up under the same conditions as described in the crystallization paper (Sabini *et al.*, 1999), but with glycine buffer instead of TRIS. Under these modified conditions, only a limited number of useful crystals was obtained. Successful removal of TRIS was checked by collection of X-ray data on one of these crystals (see below). Subsequently, a TRIS-free crystal was soaked in mother liquor containing 100 m*M* mannotriose in order to produce a sugar complex.

Table 1
Data statistics.

Figures in parentheses relate to the outer shell.

	Pty	Nat^{RT}	Nat$^{CC}_{TRIS}$	Nat^{CC}	Man2
Temperature (K)	110	300	110	110	110
Synchrotron	SRS	EMBL-HH	SRS	ESRF	SRS
Beamline	9.6	X11	9.6	ID14.4	9.6
Wavelength (Å)	0.87	0.9095	0.87	0.946	0.87
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell dimensions					
a (Å)	51.1	50.0	50.2	50.2	50.7
b (Å)	54.3	54.3	54.6	54.6	54.9
c (Å)	61.1	60.2	60.8	60.8	61.4
β (°)	110.2	111.3	111.2	111.2	111.7
V_m (Å^3 Da^{-1})	2.0	1.9	1.9	1.9	2.0
Solvent content (%)	37.6	34.9	36.1	36.1	37.6
Resolution (Å)	20.0–1.65	20.0–2.00	20.0–1.50	20.0–1.50	20.0–1.40
R_{merge} (%)	3.9 (7.1)	3.0 (3.5)	4.2 (10.4)	6.3 (27.7)	6.3 (24.0)
I/σ_I	18.4 (9.8)	42.8 (20.8)	22.4 (8.4)	20.9 (4.2)	17.3 (2.6)
Unique reflections	60810	19562	40304	48906	50979
Redundancy	2.0 (1.2)	5.6 (2.8)	2.4 (1.4)	3.7 (3.4)	3.2 (2.1)
Completeness (%)	91.0	97.8	84.1	99.8	91.2

3.2. Data collection

Native data were initially recorded at room temperature to 2.0 Å resolution (Nat^{RT}) from a crystal mounted in a glass capillary, using synchrotron radiation on beamline X11 at the EMBL Hamburg Outstation. The detector was a MAR Research imaging plate and the data were processed with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997; Table 1). Subsequently, cryo-protectant conditions were established using a solution containing the well precipitant and 25%(*v*/*v*) glycerol. All subsequent data were recorded at 110 K from vitrified crystals mounted in loops.

As none of the available coordinates of homologous proteins proved to be a satisfactory molecular-replacement (MR) model for the *T. reesei* β-mannanase, a search for heavy-atom derivatives was performed. After preliminary screening in house, data for two potential derivatives were collected using the ASDC QUANTUM-4 CCD detector on beamline 9.6 at the SRS, Daresbury Laboratory. For the crystal soaked in platinum (II) (2,2′:6′,2″ terpyridine) chloride dihydrate, data were recorded to 1.65 Å resolution and for the second crystal soaked in K_2PtCl_4, data were recorded to 1.5 Å. The wavelength was 0.87 Å, just below the platinum L_I absorption edge. 1.5 Å resolution data

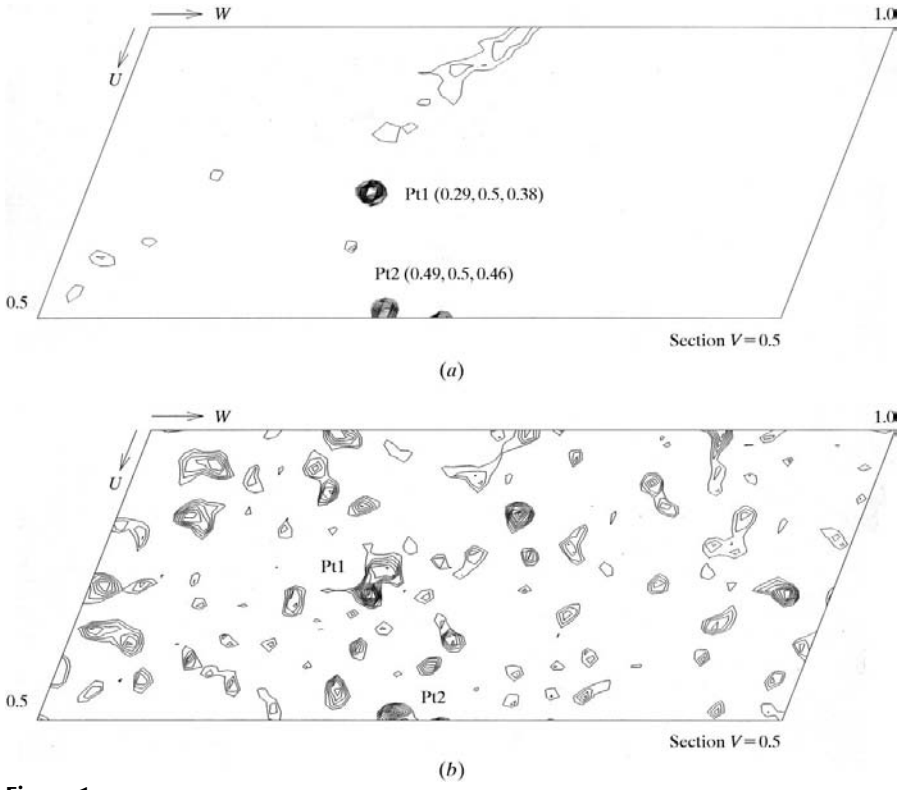


Figure 1
(*a*) Anomalous and (*b*) isomorphous difference Patterson syntheses at 1.65 Å resolution for the Pty complex: the Harker vectors of the two platinum sites are indicated.

Table 2
Phase differences for the SAD, *DM* and *REFMAC*/*ARP* phase sets from those computed from the final model.

Phase	Average difference (°)
SAD	60.6
DM	49.1
REFMAC/*ARP*	21.5

were subsequently collected at the ESRF station ID14.4 with a ASDC QUANTUM-4 CCD detector on a TRIS-free crystal in order to confirm that the TRIS had indeed been removed from the active site. Finally, 1.4 Å resolution data were collected from a TRIS-free crystal soaked overnight in a solution of 100 m*M* mannotriose at SRS station 9.6. Statistics for the five data sets are listed in Table 1.

The platinum(II) (2,2′:6′,2″ terpyridine) chloride dihydrate complex (Pty) resulted in an excellent anomalous difference Patterson synthesis (Fig. 1*a*) showing the location of two closely spaced platinum ions with peak heights of more than 29σ (at 1.65 Å resolution). In contrast, while the isomorphous difference Patterson did contain features roughly corresponding to the two heavy atoms (Fig. 1*b*), there were several other spurious features of essentially the same height on the Harker section. Indeed, the data for the Pty derivative scaled poorly to both the room-temperature native and that collected under cryogenic conditions, with mean fractional isomorphous differences of 41.3 and 48.2%, respectively, confirming significant non-isomorphism (Fig. 2). It was decided not to use the differences between native and Pty in the phase calculation, in order to avoid the systematic errors likely to be introduced. The anomalous Patterson of the K_2PtCl_4-soaked crystal was featureless and these data were used as the first cryogenic 'native' (Nat^{CC}_{TRIS}).

3.3. Phase determination for the Pty complex

Initial phases were computed for the Pty structure using only the anomalous scattering information from the two Pt atoms. The positional and displacement parameters were refined and phases were computed using *MLPHARE* (Collaborative Computational Project, Number 4, 1994; Table 2). The electron-density map computed with these phases showed significant contrast between the protein and solvent regions. The resulting map was subjected to solvent flattening using the program *DM* (Cowtan & Main, 1993). This gave substantial enhancement of the phase information and a map with good contrast and continuity within the protein region. Whilst it would have been possible to build a model into this map, an automated procedure was used to quickly build and refine a model of a 'dummy atom' pseudo-water model.

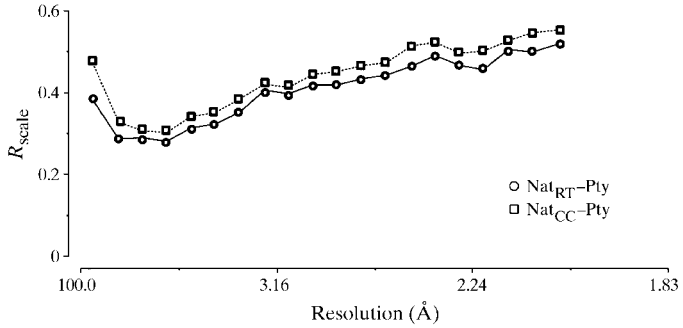


Figure 2
Mean fractional isomorphous difference for the Pty data against the native at room temperature and under cryogenic conditions.

3.4. Building and refinement of the pseudo-water model

The initial model was composed of the two Pt atoms alone. This was extended to a pseudo-water model in an automated iterative procedure consisting of two steps.

The first step involved unrestrained phased maximum-likelihood refinement in reciprocal space against observed structure-factor amplitudes and the SAD phases as observations using *REFMAC* (Murshudov *et al.*, 1997; Pannu *et al.*, 1998). The SAD phases, represented as a bimodal phase probability distribution (Hendrickson & Lattman, 1970), were

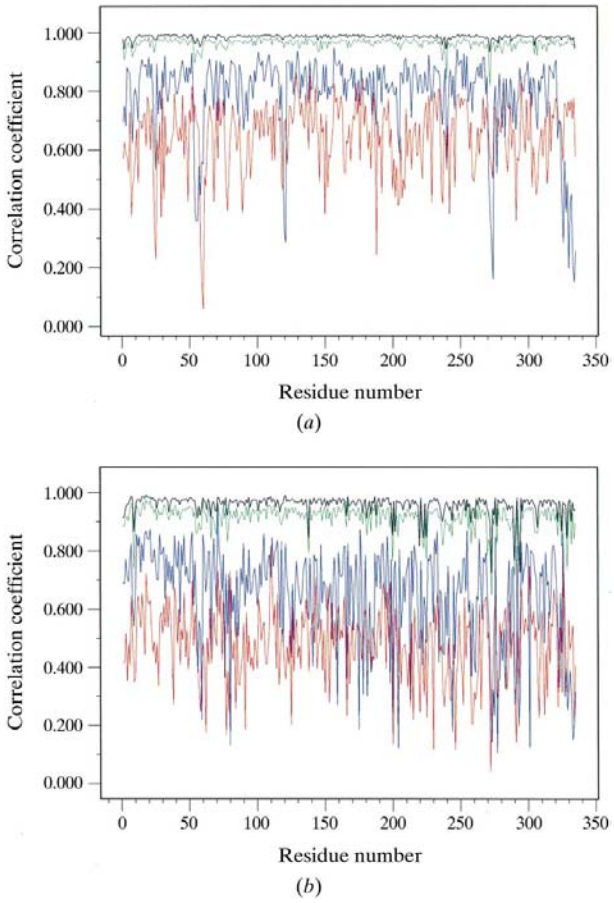


Figure 3
Map correlation for (*a*) main and (*b*) side chains, showing the quality of the calculated phases. The correlation coefficients were calculated for each residue using the *CCP*4 program *OVERLAPMAP* (Collaborative Computational Project, Number 4, 1994). The SAD phases are of good quality and, as expected, they improve after *DM*. The phases after the *REFMAC*/*ARP* procedure (without the contribution of *DM*) are essentially as good as the final phases after the anisotropic refinement.

Table 3
Refinement statistics.

	Pty	Nat^{RT}	Nat^{CC}_{TRIS}	Nat^{CC}	Man2
R factor (%)	11.5	12.6	11.1	12.7	12.4
R_{free} (%)	16.0	20.0	15.7	17.4	17.1
Ramachandran plot: non-glycine residues in					
most favoured region (%)	93.3	92.0	93.3	93.6	93.0
allowed region (%)	6.4	7.7	6.7	6.4	7.0
generously allowed region (%)	0.3	0	0	0	0
disallowed region (%)	0	0.3	0	0	0
No. of atoms/molecules					
Protein†	2673	2676	2708	2699	2707
Pt–terpy	19 × 2	—	—	—	—
NAG	14 × 4	14 × 4	14 × 4	14 × 4	14 × 4
Sulfate	5 × 2	—	5 × 1	5 × 1	5 × 1
Glycerol	—	—	6 × 3	6 × 1	6 × 2
TRIS	—	8 × 1	8 × 1	—	—
Mannobiose	—	—	—	—	22 × 1
Solvent molecules	606‡	407	458	478	497
1–2 bond deviation (Å)	0.013 (0.020)	0.015	0.013	0.016	0.015
1–3 angle distance (Å)	0.035 (0.040)	0.034	0.029	0.031	0.031
Mean *B* factor (Å^2)					
Protein	10.9	10.2	12.8	11.8	12.6
Main chain	10.3	9.4	12.1	11.1	11.8
Side chains	12.4	11.6	14.3	12.4	13.3
Pt–terpy	12.8	—	—	—	—
NAG	29.7	24.7	27.6	30.6	30.1
Sulfate ions	18.8	—	24.4	36.3	36.0
Glycerol	—	—	22.8	26.3	30.2
TRIS	—	6.3	9.3	—	—
Mannobiose	—	—	—	—	19.6§
Solvent	31.0	24.6	30.6	27.7	32.0
PDB code¶	1qnq	1qno	1qns	1qnp	1qnr
Structure-factor codes¶	r1qnqsf	r1qnosf	r1qnssf	r1qnpsf	r1qnrsf

† This apparent discrepancy reflects the different numbers of alternate conformations. ‡ The higher number of waters in the Pty complex reflects a more generous selection procedure in the *ARP* protocol. § The mannobiose was refined with half occupancy. ¶ Both the structure factors (.SF) and the coordinates (.PDB) have been deposited.

they were merged and the electron-density σ-level cutoff. A good map was obtained using the following protocol.

(i) Inclusion of the SAD phases as observations was essential. Unphased refinement failed to improve the model and resulting phases. The experimental phases were represented as a bimodal phase-probability distribution (Hendrickson & Lattman, 1970).

(ii) The 'bulk' solvent correction was applied (Murshudov *et al.*, 1997).

(iii) Atoms were added to the model if there was electron density in the F_oF_cWT map at a level higher than 3σ and if the separation from an atom already present in the model was within the range 1.0–3.3 Å. Atoms were removed from the model if their density in the $2F_oF_cWT$ map was lower than 1σ and whenever the above distance criteria were not fulfilled. Pairs of atoms closer than 1.1 Å were merged.

In the final *ARP* building of the model from the two platinum sites, up to 300 atoms were allowed to be added in each cycle and up to 300 were allowed to be removed. All data in the resolution range 20–1.65 Å were included. The procedure essentially converged, with a total of 3684 pseudo-water atoms after 60 cycles. In the initial cycles, only a small number of atoms were added, reflecting the building from only two starting atoms. From about cycle 10, the number added rapidly increased and, by cycle 35, more than 200 atoms were being added in each cycle. The final *R* factor was 14.55% and the R_{free} was 23.43%. The large difference in the two *R* factors reflects the over-fitting of the model with pseudo-water molecules without proper stereochemical restraints for a protein.

combined with those calculated from the current model in each cycle.

The second step involved updating of the current atomic model in real space using maps computed with the coefficients $2m_{comb}|F_o|\exp(i\varphi_{comb}) - D|F_c|\exp(i\varphi_{calc})$ ($2F_oF_cWT$) and $m_{comb}|F_o|\exp(i\varphi_{comb}) - D|F_c|\exp(i\varphi_{calc})$ (F_oF_cWT), where *m* is the FOM for the combined phases and *D* is a function of the coordinate error and resolution (Murshudov *et al.*, 1997). The maps were used to update the current model using *ARP* (Lamzin & Wilson, 1993) to add new atoms (O atoms) if the difference map (F_oF_cWT) showed positive features and to remove atoms when the $2F_oF_cWT$ showed no electron density for those atoms. Real-space refinement was carried out to optimize atomic positions before the first step was iterated.

Appropriate protocols for phase refinement and model generation were investigated by varying a number of parameters with *ARP* and *REFMAC*. In the unrestrained refinement with *REFMAC*, the distribution of phase probability (bimodal *versus* unimodal), the scale type (simple *versus* bulk solvent) and the scale-shift parameters were varied. In *ARP*, the variables were the number of atoms added and removed at each cycle, the distance between both new–old and new–new atoms, the minimum separation between two atoms before

3.5. Building and refining the Pty model

After the automated *REFMAC*/*ARP* procedure converged, the resulting pseudo-water model provided an extremely good $2F_oF_cWT$ electron-density map. The main chain showed essentially complete connectivity and the majority of the side chains could be identified. For example, the density for tyrosine and tryptophan side chains displayed holes in the aromatic rings and the four disulfide bridges had beautifully defined density. Building of the protein model was carried out using *QUANTA* (Molecular Simulations Inc., San Diego, USA) and led to an essentially complete protein model in a single graphics session.

The model for the entire protein chain (344 amino acids) was subjected to restrained refinement in *REFMAC*. Additional electron-density features connected to asparagines 130, 157, 250 and 238 were assigned to four sites of N-glycosylation (NAG) out of the five potential glycosylation sites. Two platinum terpyridine molecules were included, one bound to His35 and the second stacked against the first through van der Waals interactions. Two sulfate ions, one TRIS and three glycerol molecules were identified. When the *R* factor reached 25% (R_{free} = 27%), water molecules (654 in total) were added iteratively using *ARP* and the *R* factor dropped to 15.8% (R_{free} = 19.4%). The refinement was completed with a few cycles of refinement with individual anisotropic temperature factors (Murshudov *et al.*, 1999), giving a final *R* factor of 11.5% (R_{free} = 16.0%; Table 3). This drop in R_{free} clearly justified the introduction of an anisotropic model.

3.6. The native structure: Nat^{CC}_{TRIS}

A starting native model was created from the Pty coordinates by deleting the platinum(II) (2,2′:6′,2″ terpyridine) chloride dihydrate molecules, the sulfate groups and all the water molecules with a *B* factor higher than 40 Å^2. This initial model was transformed into the native unit cell using the *CCP*4 program *COORDCONV* (Collaborative Computational Project, Number 4, 1994). After rigid-body adjustment, inclusion of ligands (458 water molecules, four NAG, one sulfate group, three glycerols and one TRIS) using *REFMAC* and *ARP* and an anisotropic model, the final *R* factor was 11.1% (R_{free} = 15.7%; Table 3).

3.7. The room-temperature native Nat^{RT}, TRIS-free native Nat^{CC} and the mannobiose complex Man2 structures

These were refined with similar protocols to those used for Nat^{CC}_{TRIS}, starting from the refined Nat^{CC}_{TRIS} model, using *REFMAC* and *ARP*. The refinements are summarized in Table 3.

4. Results and discussion

The three-dimensional structure of the catalytic core domain of β-mannanase from *T. reesei* has been solved using the phase information derived from a single non-isomorphous platinum terpyridine complex (Pty). The automated model-building and refinement procedure produced excellent electron-density maps that clearly showed the presence of all 344 protein residues and the four disulfide bridges. The refined Pty structure was used as a starting point for the other four coordinate sets presented here and the high quality of the final models can be seen from the refinement statistics in Table 3.

4.1. Quality of the map

The initial map calculated using the phases from single anomalous diffraction (SAD) was acceptably good and contained sufficient information to be improved though density modification (Cowtan & Main, 1996). This improvement gave an electron-density map that was of sufficient quality such that manual model building could have begun, although substantial human input would have been required. However, the solvent-flattened phases were not utilized in phased refinement, in part to evaluate the power of automated refinement using the *REFMAC*/*ARP* procedure starting from minimal phase information. In addition, this avoided both the overestimation of the FOM often found with *DM* phases and the 'edge' effects at the molecular surface dependent on any ambiguities in the envelope definition.

The quality of the calculated phases was assessed by plotting the correlation coefficient between a particular map and

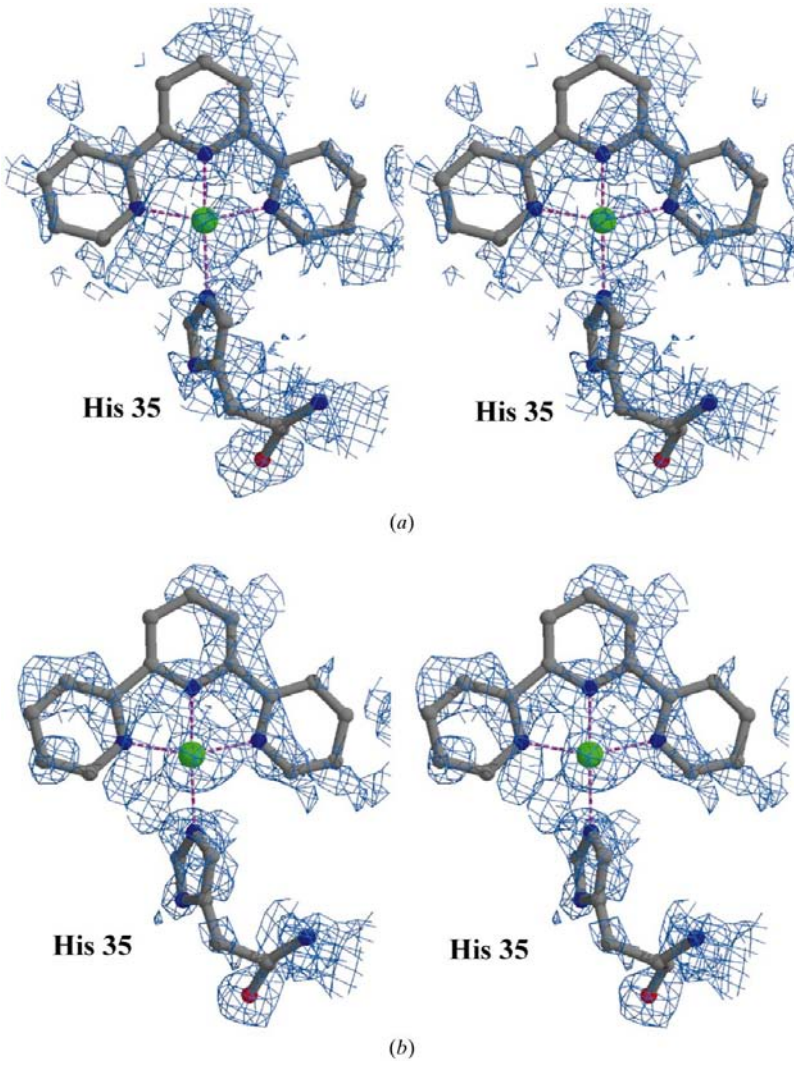


Figure 4
Density map (*a*) after SAD, (*b*) after *DM*, (*c*) after the *REFMAC*/*ARP* cyclic procedure and (*d*) after the final anisotropic refinement for the platinum (II) (2,2′:6′,2″ terpyridine) compound coordinated, through the platinum, with the NE2 atom of His35 from an adjacent symmetry-related molecule.

that computed with phases from the final model, as a function of residue number, using the program *OVERLAPMAP* (Collaborative Computational Project, Number 4, 1994; Figs. 3*a* and 3*b*). Both the SAD and the solvent-flattened maps showed several breaks in the main-chain density, which roughly correspond to regions where the map correlation is poor. The *REFMAC*/*ARP* procedure produced a map that showed high connectivity along the entire main chain and was almost as good as the map calculated with the final model phases. Figs. 4 and 5 showed the electron density in two representative regions at various stages of the phasing and refinement procedure.

Fig. 4 shows the construction of the bound ligand lying in the active site of the protein. The SAD electron-density map (Fig. 4*a*) has major perturbations around the Pt atom: these were only partially corrected by *DM* (Fig. 4*b*), but were corrected excellently by the *REFMAC*/*ARP* procedure (Fig. 4*c*).

Fig. 5 shows the density for some representative residues lying on the surface of the protein, including Trp164. The density is generally poorly defined in the SAD map (Fig. 5*a*) and only slightly better in the *DM* map (Fig. 5*b*), reflecting the limitation of density modification at the protein–solvent boundary, but was again excellent in the *REFMAC*/*ARP* map (Fig. 5*c*).

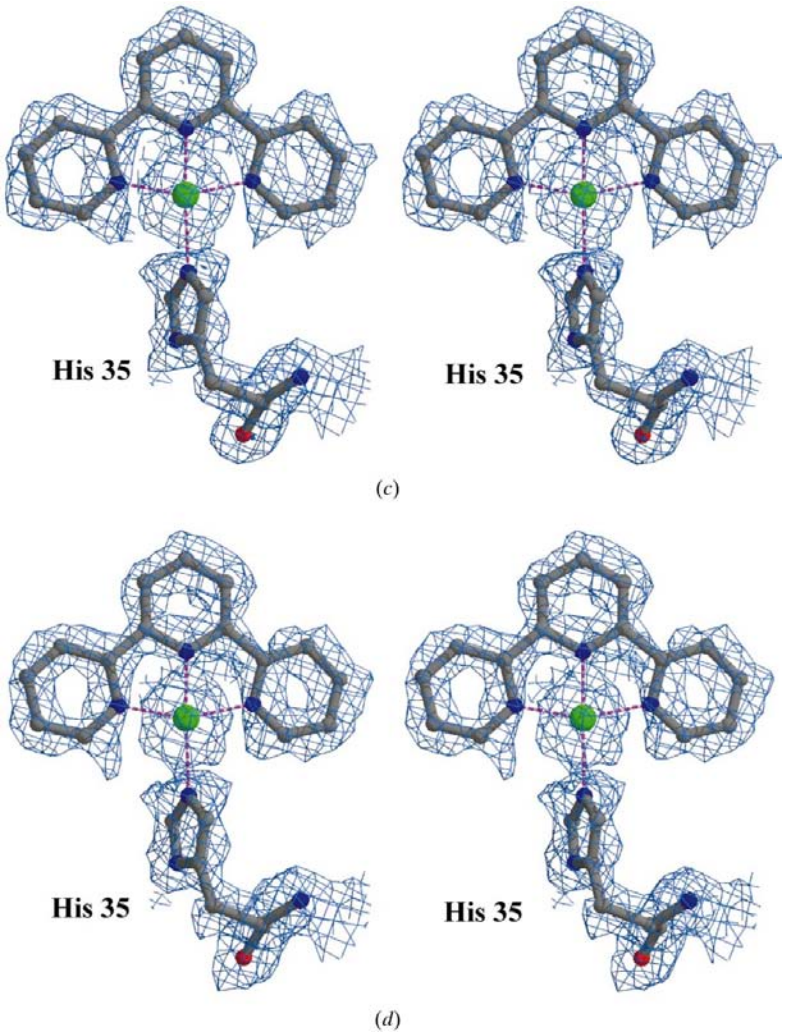


Figure 4 (continued)

4.2. Overall structure description

T. reesei β-mannanase shows the classical $(\beta\alpha)_8$-barrel architecture typical of the family 5 glycosyl hydrolases (Fig. 6). Among the 77 *O*-glycosyl hydrolases classified to date, families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 51 and 53 (relatedness based on sequence similarities only), which belong to the GH-A clan, and families 13 and 70 of the GH-H clan display a $(\beta\alpha)_8$-barrel motif.

The overall fold of the *T. reesei* β-mannanase is very similar to that of the *Therm. fusca* enzyme and will not be described in detail here. However, visual comparison of the fungal and the bacterial structures revealed several interesting differences. Both structures contain two short β-strands (β1 and β2) at the N-terminus which cover the bottom of the barrel, but the *T. reesei* β-mannanase contains two additional β-sheets. A three-stranded sheet (βA, βB and βC) is inserted between β5 and α3 and two β-strands (βD and βE) lie close to the C-terminus. Both these β-sheets extend over the putative -3 substrate subsite (see below) and could be involved in the interaction with the substrate: this implies the existence of more sites in the fungal than in the bacterial enzyme.

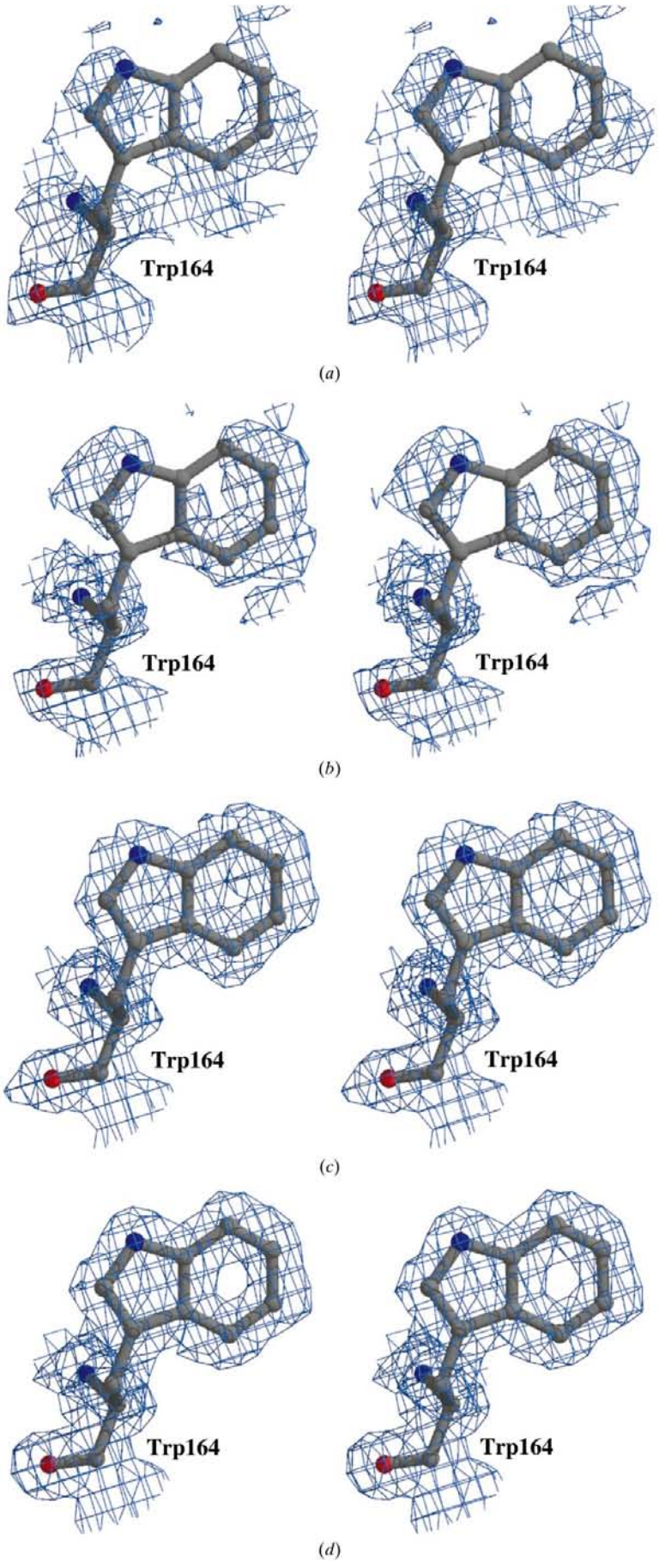
A single *N*-acetyl glucosamine residue was built at each of four glycosylation sites with well defined density. Typically, the bacterial enzyme does not show such glycosylation.

The *Therm. fusca* enzyme contains a single disulfide bond between Cys74 and Cys81, whereas the *T. reesei* mannanase contains four disulfide bonds: Cys26 and Cys29 lie on α1, Cys172 and Cys175 link across the β10–α7 coil, Cys265 and Cys272 connect α11 and coil α11–β13, and Cys284 and Cys334 link α12 and α13.

4.3. The active site

The position of the active site was clearly identified in the structural study of the homologous family 5 subgroup A8 bacterial *Therm. fusca* mannanase (Hilge *et al.*, 1998). The two catalytic glutamate residues lie in a groove across the enzyme's surface at one end of the $(\beta\alpha)_8$ barrel. Four subsites were identified in the complexes of *Therm. fusca* mannanase with mannose oligomers, with sugars being bound in the subsites -4, -3 and -2 and the position of site -1 being deduced from the position of the glutamates.

The active site of the *T. reesei* β-mannanase is in an equivalent position at one end of the $(\beta\alpha)_8$-barrel. Sequence alignment reveals that only eight residues are strictly conserved in all family 5



mannanases and cellulases (Hilge *et al.*, 1998). In *T. reesei* β-mannanase, they are Arg54, His102, Asn168, Glu169, His241, Tyr243, Glu276 and Trp306. The significance of seven of these residues is clear, as they all lie in and around the active site. Glu169 and Glu276 have the roles of catalytic acid/base and nucleophile, respectively. Arg54 is hydrogen bonded to Asn168, which is in turn hydrogen bonded to Glu169. His241 is also hydrogen bonded to Glu169. Tyr243 is hydrogen bonded to Glu276. Trp306 forms the hydrophobic sugar-binding platform in subsite −1. Only His102 lies on the opposite side of the molecule, with no obvious functional role in catalysis or substrate binding.

There is a non-prolyl *cis* peptide bond between residues Trp306 and Gln307. It is believed to be essential for the enzyme function, since it constrains the position of Trp306, which is involved in the interactions with the −1 subsite. In the *Therm. fusca* mannanase, there is an equivalent *cis* peptide bond between Trp254 and Ser255 (Hilge *et al.*, 1998).

In the *T. reesei* structures described here, a symmetry-related protein molecule occupies the putative −3 and −4 sites, preventing binding of substrate analogues to these two subsites in this crystal form. Indeed, there is possibly some conflict with access to site −2. The subsites −2, −1 and +1 contain a number of different moieties in addition to water in the present series of structures, reflecting the presence of cryoprotectant and buffer in the crystallization medium (see Table 3).

In both the Nat^{RT} and Nat^{CC}_{TRIS} structure, a well defined TRIS molecule is bound in the −1 site and forms hydrogen bonds to the two catalytic residues, Glu169 and Glu276, the conserved Asn168 and three water molecules (Fig. 7). The protein had been stored in 10 m*M* TRIS pH 7.5 and the same buffer was used in the original crystallization medium at 0.1 *M* concentration (Sabini *et al.*, 1999). TRIS is well known as a saccharide analogue in glycoside hydrolases (James & Lee, 1996; Kimura *et al.*, 1997). This TRIS molecule prevented the binding of mannose in a series of soaking experiments (data not shown). In addition, three glycerol molecules were modelled into the electron density in the Nat^{CC}_{TRIS} structure. The first is in subsite −2 and the second and third are both around subsites +1 on platforms formed by Trp114 and Tyr27.

For the Nat^{CC} crystal, the TRIS was removed from the enzyme by exchanging the buffer before crystallization and the absence of TRIS in the −1

Figure 5
As Fig. 4. Aromatic residues (here Trp164 is shown) had a 'hole' at the centre of the ring in the *REFMAC*/*ARP* map (*c*).

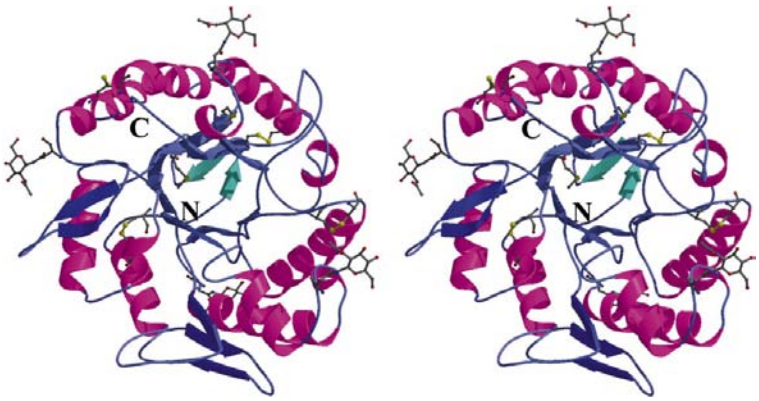


Figure 6
Secondary structure of the *T. reesei* β-mannanase. As in the *Therm. fusca* mannanase, two short β-strands (sky blue) at the N-terminus form the bottom of the barrel. In addition to these, the fungal enzyme also contains a three-stranded and a two-stranded β-sheet (both in blue) that lie close to the C-terminus. The four N-glycosylation sites and the four disulfide bridges are also shown. This figure was generated using the program *BOBSCRIPT* (Kraulis, 1991; Esnouf, 1997)

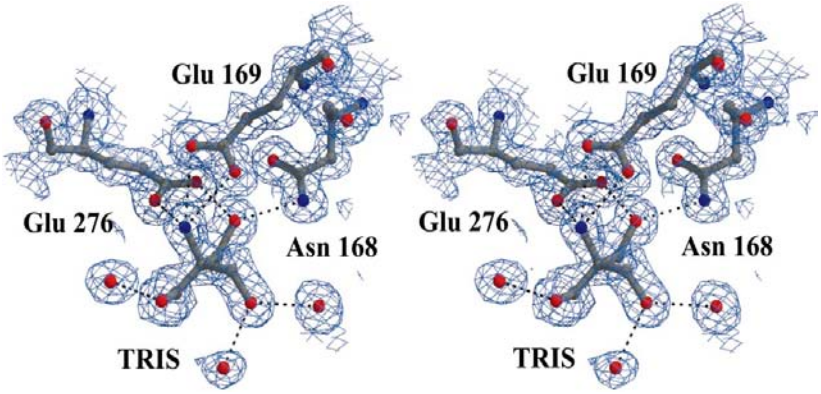


Figure 7
Active site of the *T. reesei* β-mannanase showing the TRIS molecule in the -1 site. The hydrogen bonds with the catalytic acid/base and nucleophile and with Asn168 are indicated by black dotted lines. The electron-density map is a maximum-likelihood-weighted $2F_{\rm obs} - F_c$ synthesis (royal blue) contoured at a level of 0.5 e Å^{-3} (1σ).

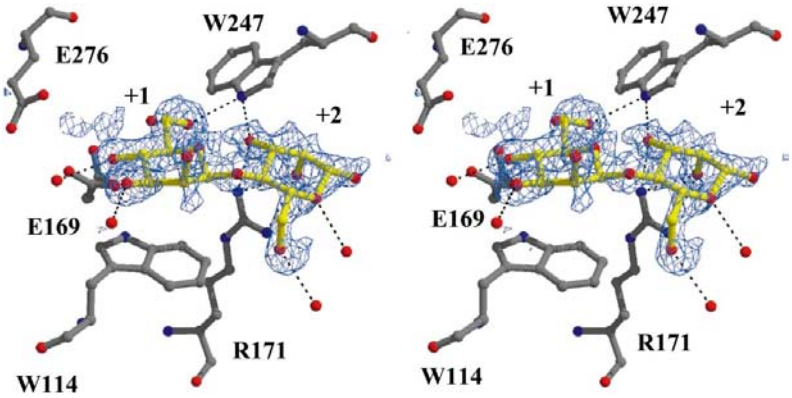


Figure 8
The maximum-likelihood-weighted $F_{\rm obs} - F_{\rm omit}$ synthesis for the mannobiose in the +1 and +2 sites contoured at 0.19 e Å^{-3} (2.8σ). The mannobiose atoms were excluded from the structure-factor calculation and several cycles of minimization were run. The hydrogen bonds (black dotted lines) are shown for the sugar (yellow) with the protein (mid grey) and the water molecules (red). Drawn using *BOBSCRIPT* (Kraulis, 1991; Esnouf, 1997).

site was confirmed from the electron density in the resulting structure. Initially, a difference map was computed using observed amplitudes from the Nat$^{\rm CC}$ minus those of the Nat$^{\rm CC}_{\rm TRIS}$ with the phases from the Nat$^{\rm CC}_{\rm TRIS}$ model, which showed negative features corresponding to the TRIS molecule. TRIS was therefore removed from the model. There was only a single glycerol in this structure in interpretable density in subsite -2.

The Man2 structure at 1.4 Å resolution from a TRIS-free crystal soaked overnight in the mannotriose solution revealed the presence of a mannobiose moiety in subsites +1 and +2. The *B* values suggested the sites were only partially occupied as when refined with full occupancy, the *B* values for the ligand were about 34 Å^2 compared with about 16 Å^2 for the surrounding protein. Therefore, an occupancy of 50% was assigned. Fig. 8 shows the electron density for the disaccharide and the residues involved in direct contacts with it and the catalytic acid/base and nucleophile. In this structure, two glycerol residues are bound in subsites -2 and -1. The +1 and +2 subsites were not identified in the *Therm. fusca* mannanase.

The mannosyl residue in the +1 site is bound through hydrophobic stacking with Trp114, which plays the role of a hydrophobic platform. The OH(3), OH(4) and OH(6) are at hydrogen-bonding distances to water molecules and, in addition, the OH(6) H atom binds the N atom of Trp247. In fact, the OH(2) bound to the C atom is not stabilized by any hydrogen bonds.

The D-mannose in the +2 subsite interacts with both Arg171 and Trp247 and two water molecules. Arg171 plays a crucial role in the positioning of the +2 ring and hence of the whole disaccharide. The *Therm. fusca* mannanase lacks both this residue and Trp247 (see amino-acid sequence alignment in Fig. 9). NH1 and NH2 of Arg171 make hydrogen bonds with the axial OH(2) hydroxyl group in the +2 subsite. In addition, the NH2 atom interacts with both the OE2 of Glu205 and the OH(3) of the sugar in subsite +2, which simultaneously hydrogen bonds to the N atom of Trp247. The OE1 of Glu205 is at a hydrogen-bonding distance to the NE of Arg171. Glu205 is not shown in Fig. 8. The -1 and -2 subsites are occupied by two glycerol molecules in this complex.

The +1 and +2 sugar residues are both well defined, with their orientation clearly established. Since the crystal was soaked in mannotriose, this suggests one of three possibilities. The first is that mannotriose is bound, but that

the third residue is disordered and its position is not visible in the density. However, there is really no density extending from either the reducing or non-reducing end to indicate the presence of an extra residue and this would suggest this possibility is unlikely. The second possibility is that there is some mannobiose in the mannotriose sample used. The third, and we believe the most likely, possibility is that the mannotriose has been hydrolysed to mannobiose in the crystal. This would suggest that *T. reesei* β-mannanase is active on mannotriose and that the complex represents a product complex and would be in keeping with results of kinetic measurements which showed low activity of the enzyme on mannotriose (Harjunpää *et al.*, 1995, 1999). This would require mannotriose to initially bind either in sites −2, −1, +1 or −1, +1, +2. The former is more probable given the likely need to distort the ring is subsite −1 before catalysis can occur. The formation of a product complex revealing subsites +1 and +2 may indeed be a fortuitous result based on the blockage of sites −3 and −4 in this crystal form.

In the platinum terpyridine complex, the active site is occupied by two molecules of platinum (II) (2,2′:6′,2″ terpyridine) chloride dihydrate: one Pt atom is coordinated through a bond with the NE2 atom of His35 from a symmetry-related molecule, while the other molecule lies within van der Waals

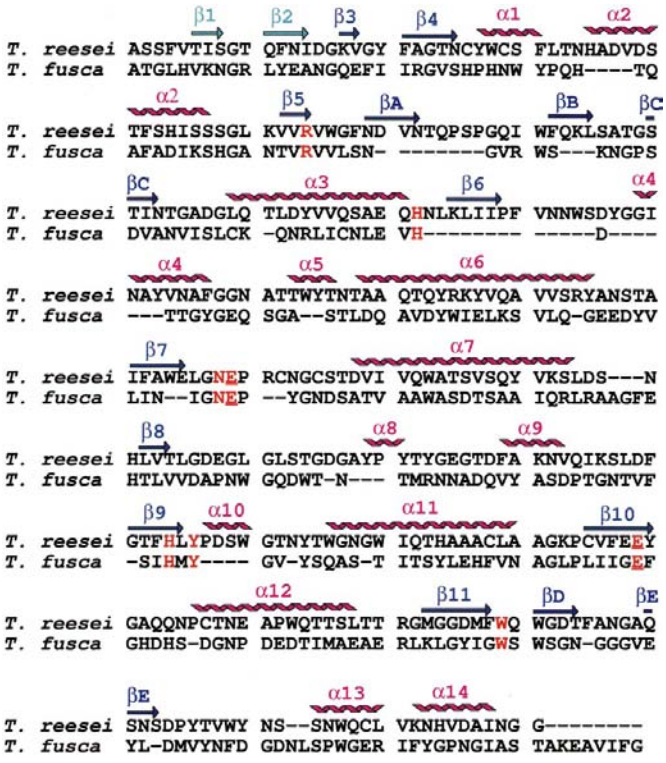


Figure 9
Amino-acid sequence alignment of *T. reesei* and *Therm. fusca* β-mannanases. The eight residues which mannanases within family 5 share with more than 60 cellulases are indicated in red. The catalytic residues of the two mannanases are underlined. The secondary structure for the fungal enzyme is indicated. Helices are in magenta, strands β3 to β11 are typical for a $(\beta\alpha)_8$ barrel (in lilac), β1 and β2 lie at the N-terminus (sky blue) and βA to βE (in blue) are present in *T. reesei* mannanase, but not in *Therm. fusca*.

contact, parallel and upside down with respect to the first one. This second aromatic molecule interacts with Trp114 within the active site. Since this derivative was obtained by diffusing the platinum compound into pregrown crystals, this result suggests that the compound binds strongly enough to displace the TRIS from the active site and to perturb the relative conformation of the protein in the crystal without disrupting the order of the lattice.

5. Conclusions

The structure was solved from the anomalous scattering information alone of the two Pt atoms in a rather non-isomorphous complex. The resulting map was of sufficient quality for an essentially complete pseudo-atom model to be constructed with the help of atom selection through *ARP* and phased maximum-likelihood refinement with *REFMAC*. The protocol is subject to considerable enhancement using recent developments. Firstly, the use of advanced phasing algorithms such as *SHARP* (La Fortelle & Bricogne, 1997) would allow the incorporation of, for example, the weak isomorphous phase information without the risk of bias. Secondly, recent developments in the *ARP* program or other similar suites could certainly be used to build a more extensive starting model into the SAD map, rather than merely using the two Pt atoms. This should considerably enhance the rate of convergence. Thirdly, the solvent-flattened map could be used to construct a better starting model and to provide more accurate phases to be used as observables in *REFMAC*, even if these are relaxed to the SAD phases to avoid bias as the model develops.

This procedure should prove to be applicable to other proteins, for instance using a well measured Se-Met derivative at a single wavelength below the selenium edge. The enhancements suggested above should allow extension of the use of the method to much poorer starting phases.

T. reesei β-mannanase showed the expected similarity to the structure of the *Therm. fusca* enzyme. The *T. reesei* enzyme shows four N-glycosylation sites. The overall fold of the catalytic domain was a $(\beta\alpha)_8$ barrel, with differences between the structures primarily in the external loops on the surface of the molecule. The active site itself is highly conserved, but the extra secondary structural elements of the fungal enzyme may provide extra subsites compared with the bacterial one. In the present structure, sites −4 and −3 (and possibly −2) are blocked by a symmetry-related protein molecule. However, after removal of the TRIS from the system, the mannobiose complex showed binding of a disaccharide in subsites +1 and +2, which was not identified from the *Therm. fusca* structures. The structure suggested that the mannotriose used in the soaking experiment had been hydrolysed to mannobiose in the crystals, confirming kinetic evidence that the enzyme is active against mannotriose.

The studies confirm the well established phenomenon of inhibition of glycosyl hydrolases by TRIS. The TRIS was bound sufficiently strongly to resist its displacement when soaking the crystals in mannotriose [although it was displaced

by the platinum (II) (2,2′:6′,2″ terpyridine) chloride in forming the heavy-atom complex]. We also observed binding of the glycerol cryoproctectant in sugar subsites −2, −1 and +1. Although the successful formation of the Man2 complex suggests it was weakly bound and readily displaced by mannobiose, this indicates that glycerol can behave as a sugar analogue and for some systems might require the establishment of alternative cryoprotectant conditions [glycerol was used as cryoprotectant at a concentration of 25%(*v*/*v*)].

Most importantly, the blockage of sites −3 and −4 in the present crystal form requires an alternative crystal form to be found for investigation of substrate binding in this site with the present enzyme, perhaps involving co-crystallization with oligosaccharide. The coordinates of the *Therm. fusca* enzyme were already kindly provided to us by Hilge and colleagues and a detailed comparison of the structures and their complexes, together with those of the other enzymes from the same family, will be published later.

The authors would like to thank the BBSRC for providing infrastructure funding through provision of a Structural Biology Centre award. KSW acknowledges support for ES through the EC BIOTECH contract BIO4-CT97-2303 (York). We thank the EC for the TMR/LSF CT98-01334 (EMBL Hamburg Outstation) contract for data collection in Hamburg and the ESRF and CCLRC for support of data collection at ESRF and SRS Daresbury, respectively. We thank the CCP4 for funding for GNM. Professor Marc Clayessens is thanked for providing mannotriose.

References

Béguin, P. (1990). *Annu. Rev. Microbiol.* **44**, 219–248.
Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Buchert, J., Salminem, M., Siika-aho, M., Ranua, M. & Viikari, L. (1993). *Holzforschung*, **47**, 473–478.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.
Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.
Davies, G. J., Dauter, M., Brzozowski, A. M., Bjørnvad, M. E., Andersen, K. V. & Schülein, M. (1998). *Biochemistry*, **37**, 1926–1932.
Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem. J.* **321**, 557–559.
Dominguez, R. & Alzari, P. M. (1995). *Nature Struct. Biol.* **2**, 569–576.
Dominguez, R., Souchon, H., Lascombe, M. B. & Alzari, P. M. (1996). *J. Mol. Biol.* **257**, 1042–1051.
Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Fierobe, H. P., Belaich, J. P., Davies, G. J. & Haser, R. (1995). *Structure*, **3**, 939–949.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 133–138.
Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. Jr & Warren, R. A. J. (1991). *Microbiol. Rev.* **55**, 303–315.
Harjunpää, V., Helin, J., Koivula A., Siika-aho, M. & Drakenberg, T. (1999). *FEBS Lett.* **443**, 149–153.
Harjunpää, V., Teleman, A., Siika-aho, M. & Drakenberg, T. (1995). *Eur. J. Biochem.* **234**, 278–283.
Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* B**26**, 136–143.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Henrissat, B. (1991). *Biochem. J.* **280**, 309–306.
Hilge, M., Gloor, S. M., Rypniewski, W., Sauer, O., Heightman, T., Zimmermann, W., Winterhalter, K. & Piontek, K. (1998). *Structure*, **6**, 1433–1444.
James, J. A. & Lee, B. H. (1996). *Biotechnol. Lett.* **18**, 1401–1406.
Kimura, A., Takata, M., Fukushi, Y., Mori, H., Matsui, H. & Chiba, S. (1997). *Biosci. Biotechnol. Biochem.* **61**, 1091–1098.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Lo Leggio, L., Parry, N. J., van Beeumen, J., Claeyssens, M., Bhat, M. K. & Pickersgill, R. W. (1997). *Acta Cryst.* D**53**, 599–604.
McCleary, B. V. (1991). *Enzymes in Biomass Conversion*, edited by G. F. Leatham & M. E. Himmel, pp. 437–449. Washington, DC: American Chemical Society.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
North, A. C. T. (1965). *Acta Cryst.* **18**, 212–216.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst.* D**54**, 1285–1294.
Sabini, E., Brzozowski, A. M., Dauter, M., Davies, G. J., Wilson, S. W., Paloheimo, M., Suominen, P., Siika-aho, M. & Penttilä, M. (1999). *Acta Cryst.* D**55**, 1058–1060.
Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R. & Karplus, P. A. (1996). *Biochemistry*, **35**, 10648–10660.
Schülein, M., Tikhomirov, D. F. & Shov, C. (1993). *Proceedings of the Second Tricel Symposium on Trichoderma Cellulases and Other Hydrolases*, Vol. 8, edited by P. Suominen & T. Reinikainen, pp. 109–116. Helsinki: Foundation for Biotechnical and Industrial Fermentation Research.
Stålbrand, H., Saloheimo, A., Vehmaanperä, J., Henrissat, B. & Penttilä, M. (1995). *Appl. Environm. Microbiol.* **61**, 1090–1097.
Stålbrand, H., Siika-aho, M., Tenkanen, M. & Viikari, L. (1993). *J. Biotechnol.* **29**, 229–242.
Viikari, L., Kantelinen, A., Siika-aho, M., Tenkanem, M., Buchert, J., Bailey, J., Pere, J. & Linko, M. (1993). *Proceedings of the Second Tricel Symposium on Trichoderma Cellulases and Other Hydrolases*, Vol. 8, edited by P. Suominen & T. Reinikainen, pp. 255–262. Helsinki: Foundation for Biotechnical and Industrial Fermentation Research.
Viikari, L., Ranua, M., Kantelinen, A., Linko, M. & Sundquist, J. (1987). *Proceedings of the Fourth International Symposium on Wood and Pulping Chemistry*, Vol. 1, pp. 151–154.
Viikari, L., Tenkanen, M., Buchert, J., Rättö, M., Bailey, M., Siika-aho, M. & Linko, M. (1993). *Bioconversion of Forest and Agricultural Plant Residues*, edited by J. N. Saddler, pp. 131–182. Wallingford: CAB International.