Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Determination of the structure of an endoglucanase from *Aspergillus niger* and its mode of inhibition by palladium chloride

Shahram Khademi,[a] Dachuan Zhang,[a]† Stanley M. Swanson,[a] Arnold Wartenberg,[b] Klaus Witte[b] and Edgar F. Meyer[a]*

[a]Biographics Laboratory, Texas A&M University, Department of Biochemistry and Biophysics, College Station, TX 77843-2128, USA, and [b]Fachrichtung Mikrobiologie, Universitat des Saarlandes, D-66041 Saarbrucken, Germany

† Current address: GeneLogic, 2001 Center Street, Suite 600, Berkeley, CA 94704, USA.

Correspondence e-mail: e-meyer@tamu.edu

Received 8 October 2001
Accepted 19 February 2002

PDB References: endoglucanase–palladium complex, 1ks4, r1ks4sf; endoglucanase, 1ks5, r1ks5sf.

The fungus *Aspergillus niger* is a main source of industrial cellulase. β-1,4-Endoglucanase is the major component of cellulase from *A. niger*. In spite of widespread applications, little is known about the structure of this enzyme. Here, the structure of β-1,4-endoglucanase from *A. niger* (EglA) was determined at 2.1 Å resolution. Although there is a low sequence identity between EglA and CelB2, another member of family 12, the three-dimensional structures of their core regions are quite similar. The structural differences are mostly found in the loop regions, where CelB2 has an extra β-sheet (β-sheet *C*) at the non-reducing end of the binding cleft of the native enzyme. Incubation of EglA with $PdCl_2$ irreversibly inhibits the EglA activity. Structural studies of the enzyme–palladium complex show that three Pd^{2+} ions bind to each EglA molecule. One of the Pd^{2+} ions forms a coordinate covalent bond with Met118 S^{δ} and the nucleophilic Glu116 $O^{\varepsilon1}$ at the active site of the enzyme. The other two Pd^{2+} ions bind on the surface of the protein. Binding of Pd^{2+} ions to EglA does not change the general conformation of the backbone of the protein significantly. Based on this structural study, one can conclude that the palladium ion directly binds to and blocks the active site of EglA and thus inactivates the enzyme.

1. Introduction

Cellulose is a linear polymer of glucosyl units connected by β-1,4 linkages. These linear chains vary in length and often consist of many thousands of units. Much of the organic material on the surface of the earth is stored in the form of cellulose. There is an enormous turnover of cellulose within the biosphere owing to large-scale production and degradation. Because of energy crises, cellulose is an important alternative source of renewable energy (Bhat & Bhat, 1997; Wyman, 1999; Lynd *et al.*, 1999). Many bacteria and fungi have developed mechanisms to break down this complex substrate, utilizing multiple enzymes collectively called cellulases. Acting together, these enzymes convert an insoluble polymer to its monomer, glucose. Current commercial applications of cellulases are mainly limited to the detergent and textile industries (Schulein *et al.*, 1998); however, present environmental concerns have refueled interest in the application of cellulases in biomass conversion (Sheehan & Himmel, 1999). For instance, using cellulases in recycling and de-inking copier-printed and laser-printed papers avoids the alkaline environment commonly required in traditional de-inking, consequently cutting chemical processing cost and significantly reducing environmental contamination (Gubitz *et al.*, 1998).

Aspergillus niger is an important commercial source of inexpensive cellulase, not only in the food and textile industries but also in pharmaceutical industries (Bakalova *et al.*, 1996). Endo-β-1,4-glucanase is the main component for cellulose degradation by *A. niger* and has been assigned to glycosyl hydrolase family 12 (Fig. 1). It is known to catalyze glycosidic bond cleavage with net retention of anomeric configuration (Henrissat & Bairoch, 1997; Zechel *et al.*, 1998). Family 12 enzymes employ a stepwise catalytic mechanism in which a glycosyl-enzyme intermediate is formed and subsequently hydrolyzed *via* oxocarbenium-ion transition states (Davies *et al.*, 1998). Analysis of hydrophobic clusters shows a significant similarity between xylanases from family 11 and endoglucanases from family 12. Therefore, these two families are considered as clan GH-C (Torronen *et al.*, 1993). In 1997, Sulzenbacher and coworkers determined the structure of an endo-β-1,4-glucanase from *Streptomyces lividans* (CelB2), a member of glycoside hydrolase family 12 (PDB entry 1nlr; Sulzenbacher *et al.*, 1997), and showed that the structure of the core region of CelB2 is very similar to that of family 11 enzymes and that the differences are in the loop regions. Here, we describe the three-dimensional structure of the catalytic domain of *A. niger* endoglucanase (EglA) as an important member of family 12.

```
S. lividans    -QADTTICEPFGTTTIQG-RYVVQNNRWG---STAPQCVTA-TDTG-FRVTQADGSAPTNG  54
S. rochei      -QADTTICEEFGSTVIQG-RYVVQNNRWG---TSATQCVTA-TDSG-FRVTQADGSVPTNG  54
R. marinus     -EPTVELCGRWDARDVAGGRYRVINNVWG---AETAQCIEVGLETGNFTITRADHDNGNN-  56
A. kawachii    ----QTMCSQYDSASSPP--YSVNQNLWGEYQGTGSQCVYVDKLSSSGASWHTKWTWSGGE  55
A. niger       -----TMCSQYDSASSPP--YSVNQNLWGEYQGTGSQCVYVDKLSSSGASWHTEWTWSGGE  54
A. oryzae      ----QELCAQYDSASSPP--YSVNNNLWGQDSGTGSQCVYVDNLSSSGAAWHTTWTWNGGE  55
A. aculeatus   --QQAQLCDQYATYTGGV--YTINNNLWGKDAGSGSQCTTVNSASSAGTSWSTKWNWSGGE  57
T. reesei      ----QTSCDQWATFTGNG--YTVSNNLWGASAGSGFGCVTAVSLSG-GASWHADWQWSGGQ  54
S. coelicolor QAAVWNSCDQWGNTSLNG--YTLYNNIWG--SGAGSQCVWANSGTDWG-----VWADHPNT  52
E. carotovora ----ASSSNDADKLYFGNNKYYLFNNVWGKDEIKGWQQTIFYNSP---ISMGWNWHWPSST  54

S. lividans    APKSYPSVFNGCHYTNCSPGTDL--PVRLDTVSAAPSSISYGFVDG-AVYNASYDIWLDPT 112
S. rochei      APKSYPSVFNGCHYTNCSPGTAL--PARISGISSAPSSISYGFVDN-AVYNASYDIWLDPT 112
R. marinus     -VAAYPAIYFGCHWAPARAIRDC--AARAGAVRRAH-ELDVTPITT-GRWNAAYDIWFSPV 112
A. kawachii    GTVKS--------YSNSGLTFD---KKLVSDVSSIPTSVTWSQDDTNVQADVSYDLFTAAN 105
A. niger       GTVKS--------YSNSGVTFN---KKLVSDVSSIPTSVEWKQDNTNVNADVAYDLFTAAN 104
A. oryzae      GSVKS--------YSNSAVTFD---KKLVSDVQSIPTDVEWSQDNTNVNADVAYDLFTAAD 105
A. aculeatus   NSVKS--------YANSGLTFN---KKLVSQISQIPTTARWSYDNTGIRADVAYDLFTAAD 107
T. reesei      NNVKS--------YQNSQIAIPQ--KRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAAD 105
S. coelicolor DGIKS--------YPNAKKVIN----KPITSLSSLTSSYNVTVPSS-GAYNTSYDIWDTD-  99
E. carotovora HSVKAYPSLVSGWHWTAGYTENSGLPIQLSSNKSITSNVTYSIKAT-GTYNAAYDIWFHTT 114

S. lividans    ARTDG--VNQTEIMIWFNRVGPIQPIGSPVGTASVGGRTWEVWSG-----GNGSNDVLSFV 166
S. rochei      PRTDG--VNRTEIMIWFNRVGQIQPIGSQVGTASVAGRTWEVWSG-----GNGTNDVLSFV 166
R. marinus     TNSGNGYSGGAELMIWLNWNGGVMPGGSRVATVELAGATWEVWYA-----DWDWNYIAYRR 178
A. kawachii    ADHAT-SSGDYELMIWLARYGSVQPIGKQIATATVGGKSWEVWYGTS-TQAGAEQKTYSFV 164
A. niger       VDHAT-SSGDYELMIWLARYGNIQPIGKQIATATVGGKSWEVWYGST-TQAGAEQRTYSFV 163
A. oryzae      QNHVT-YSGDYELMIWLARYGTIQPIGTQIDTATVEGHTWELWYGTT-IQAGAEQKTYSFV 164
A. aculeatus   INHVT-WSGDYELMIWLARYGGVQPIGSQIATATVDGQTWELWYG-----ANGSQKTYSFV 162
T. reesei      PNHVT-YSGDYELMIWLGKYGDIGPIGSSQGTVNVGGQSWTLYYG-----YNGAMQVYSFV 160
S. coelicolor --------YDYEIMLWVNHHGAVGPLGTFQGSVGLGGHTWDVYKG-----NNGANEVFSFL 147
E. carotovora DKANWDSSPTDELMIWLNDTN-AGPAGDYIETVFLGDSSWNVFKGWINADNGGGWNVFSFV 174
                         *
S. lividans    APSAISGWSFDVMDFVRATVAR-GLAENDWYLTSVQAGFEPWQN-G-AGLAVNSFSSTVE  224
S. rochei      APSAMSSWSFDVMDFVRATVAR-GLAGNDWYLTSIQAGFEPWQN-G-AGLAVNSFSSTVN  224
R. marinus     TTPTTSVSELDLKAFIDDAVAR-GYIRPEWYLHAVETGFELWEG-G-AGLRTADFSVTVQ  236
A. kawachii    AGSPINSWSGDIKDFFNYLTQNQGFPASSQHLITLQCGTEPFTG-GPATFTVDNWTASVN  224
A. niger       SESPINSYSGDINAFFSYLTQNQGFPASSQYLINLQFGTEAFTG-GPATFTVDNWTASVN  223
A. oryzae      SATPINTFGGDIKKFFDYITSKHSFPASAQYLINMQFGTEPFTG-GPVTFTVPNWTASVN  224
A. aculeatus   APTPITSFQGDVNDFFKYLTQNHGFPASSQYLITLQFGTEPFTG-GPATLSVSNWSASVQ  222
T. reesei      AQTNTTNYSGDVKNFFNYLRDNKGYNAAGQYVLSYQFGTEPFTG--SGTLNVASWTASIN  219
S. coelicolor RTSDSNSGTVNILPILKWIKDTKGWMG-NETIGDVQFGYEITSSAGGLDFRTNNLTVSGG  207
E. carotovora HTSGTNSASLNIRHFTDYLVQTKQWMSDEKYISSVEFGTEIFGG--DGQIDITEWRVDVK  233
                                                  *
```

Figure 1
Multiple sequence alignment among catalytic domain sequence of glycosyl hydrolase family 12. The conserved residues are in bold. The putative catalytic residues are marked by a star. The sequences are as follows; *S. lividance*, *Streptomyces lividans* (U04629); *S. rochei*, *Streptomyces rochei* (X73953); *R. marinus*, *Rhodothermus marinus* (U72637); *A. kawachii*, *Aspergillus kawachii* (Q12679); *A. niger*, *Aspergillus niger* (AJ224452); *A. oryzae*, *Aspergillus oryzae* (D83731); *A. aculeatus*, *Aspergillus aculeatus* (P22669); *T. reesei*, *Trichoderma reesei* (AB003694); *S. coelicolor*, *Streptomyces coelicolor* (AL355913); *E. carotovora*, *Erwinia carotovora* (P16630).

Palladium complexes have been shown to inhibit certain cellulases from various sources. For example, palladium compounds irreversibly inhibit cellobiohydrolase I and endoglucanase II, but not cellobiohydrolase II from *Trichoderma reesei* (Lassig *et al.*, 1995; Shultz *et al.*, 1995). Other metals, such as platinum and rhodium, do not inhibit these cellulases. The mode of inhibition of cellulases by palladium complexes has been a subject for conjecture (Lassig *et al.*, 1995; Shultz *et al.*, 1995). Here, we not only show the inhibition of EglA by palladium chloride, but also describe the mode of inhibition of EglA by palladium based on the structure of the EglA–palladium complex.

2. Materials and methods

2.1. Cellulase purification

A. niger (strain CBS 554.65; ATCC 16888) was cultivated in 1 l culture flasks with 300 ml of Norkans medium modified after Eriksson & Pettersson (1975). As a carbon source, 10 g of 'Profix Cellulose Flies' were added to the medium. The flasks, inoculated with conidia of *A. niger*, were incubated for 18 d at 303 K. The culture solution was separated from mycelium and cellulose by vacuum-filtration funnels. The extracted solution was concentrated 100-fold with a Minitan-Tangential 10 kDa NMWL filtration system (Millipore). The solution was subsequently filtered with a Millipore 0.45 µm membrane filter, type HAWP.

The initial gel filtration was carried out with a 1000 × 50 mm Sephacryl S200 HR column (Pharmacia) using a buffer system of 0.05 *M* KH_2PO_4 pH 6.0 and 0.15 *M* NaCl. The volume of the fifth peak was concentrated and prepared by buffer exchange for chromatofocusing performed with a 50 × 10 mm column filled with Polybuffer Exchanger 9-4 (Pharmacia). The starting buffer was 0.015 *M* 2-methylpiperazine pH 5.0. The limit buffer was Servalyt 3–6 diluted to a concentration of 0.2%(*w*/*v*), pH 3.1. The resultant enzyme solution was further purified by anion-exchange chromato-

graphy, using a 100 × 20 mm column with DEAE Si 300 (Serva Novex). The protein was eluted by applying a gradient from 10 to 100 m*M* KH_2PO_4 pH 7.4. The volume of the main peak was concentrated with a CX 10 Millipore cartridge, 10 kDa NMWL, de-salted with Sephadex G25 (Pharmacia) and finally lyophilized.

2.2. Enzyme assays

The endoglucanase activity of the NtEgl was measured by incubating a solution of 1 µg ml^{-1} EglA with 1%(*w*/*v*) sodium carboxymethyl cellulose [CMC; average molecular weight, 250 000 Da; degree of carboxymethyl substitution, 0.7(*w*/*v*); Aldrich] in 50 m*M* sodium acetate pH 5.0 at 310 K for 30 min. In order to determine the produced reducing sugar, 10 µl of the reaction solution was added to 1 ml of tetrazolium blue reagent solution (Jue & Lipke, 1985) and the mixture was then boiled for 3 min. After cooling, absorbance was determined at 660 nm. In order to determine inhibition of EglA by $PdCl_2$, a solution of 1 µg ml^{-1} EglA was preincubated with 100 µ*M* $PdCl_2$ for 15 min, after which the reaction was started by addition of 1% CMC. Degradation of CMC was linear for at least 60 min.

2.3. Crystallization and data collection

The hanging-drop vapor-diffusion method was used for crystallization of EglA. On a cover slide, 3 µl of protein solution (9 mg ml^{-1} EglA in buffer) and 3 µl of reservoir (28% PEG 4K and 100 m*M* acetate pH 4.8) were mixed and equilibrated against a 1 ml reservoir. During two weeks, the crystal was transferred twice into fresh protein solution. The size of crystal before it was transferred into fresh protein solution was about 0.1 × 0.1 × 0.1 mm. The fresh protein solution was pre-equilibrated with the buffer for about 30 min. After transferring into new protein solution, the crystal was equilibrated against the new reservoir solution. After two or three weeks, when the crystal size was 0.4 × 0.4 × 0.4 mm, the crystal was harvested. The space group of the crystal is *I*422, with unit-cell parameters $a = b = 131.77$, $c = 71.72$ Å, $\alpha = \beta = \gamma = 90^\circ$.

Data collection was performed at room temperature using a Nonius/MacScience DIP2030-k image plate and a Rigaku rotating-anode X-ray source with Cu $K\alpha$ radiation focused by an MSC/Osmic mirror system. Diffraction data were evaluated with *DENZO* and scaled using *SCALEPACK* (Otwinowski & Minor, 1997) (Table 1).

2.4. Structure determination and refinement

Patterson search methods using the *CNS* package were used to solve the phasing problem. We used the fast direct rotation-search method (DeLano & Brünger, 1995; Brünger, 1997) with the *CNS* package, using the complete CelB2 structure as the search model. Sequence comparison between CelB2 and EglA shows only 26% sequence identity (Fig. 1). Native EglA diffraction data between $d = 15$ and 4.0 Å with $F_{obs} > 2\sigma$ were used in a direct rotation-search method. The top ten peaks from the rotation search were each subjected to 30 steps of Patterson correlation (PC) refinement (Brünger, 1990) using a target of squared normalized structure factors. A fast translation search (Navaza & Vernoslova, 1995) was carried out for each of the top ten rotation-function peaks after PC refinement. The highest translation-search peaks were further improved by 30 steps of PC refinement.

Table 1
Structure determination summary for EglA.

Values for the highest resolution shells are in parentheses.

	Native EglA	EglA–palladium complex
Space group	*I*422	*I*422
Unit-cell parameters (Å, °)	$a = b = 131.57$, $c = 71.64$, $\alpha = \beta = \gamma = 90.00$	$a = b = 131.09$, $c = 71.99$, $\alpha = \beta = \gamma = 90.00$
Resolution range (Å)	30.0–2.1 (2.18–2.10)	30.0–2.5 (2.55–2.50)
Completeness (%)	98.0 (95.5)	96.0 (89.3)
Total reflections	325097	117527
Unique reflections	18660	11084
$I/\sigma(I)$	31.0 (6.9)	26.7 (3.7)
R_{merge}†	0.07 (0.30)	0.09 (0.63)
No. of protein atoms	1717	1717
No. of water molecules	124	34
No. of Pd^{2+} ions	—	3
R value‡ (%)	17.5	20.8
R_{free} (%)	19.6	24.2
R.m.d.s. bonds (Å)	0.006	0.009
R.m.d.s. angles (°)	1.34	1.43

† $R_{merge} = \sum_h \sum_i |I(h)_i - \langle I(h)\rangle| / \sum_h \sum_i \langle I(h)_i\rangle$, where *I* is the observed intensity and $\langle I\rangle$ is the average intensity of multiple observations of symmetry-related reflections. ‡ *R* value = $\sum ||F_o| - |F_c|| / \sum |F_o|$.

The refinement was performed with the maximum-likelihood method based on amplitudes (Pannu & Read, 1996). Only 90% of the data in the resolution range 30–2.1 Å were used for the refinement and the remaining 10% were reserved throughout as the test set for structure validation (σ_F cutoff = 0). The best solution after rigid-body refinement had an *R* value of 52.5% (R_{free} = 52.2%). Simulated annealing using torsion-angle molecular dynamics employed a protocol used by Adams *et al.* (1997), with initial minimization followed by torsion-angle molecular dynamics and slow cooling from 5000 to 0 K in 50 K steps. This first round of simulated-annealing refinement reduced the *R* value from 52.5 to 42.9% (R_{free} fell from 52.2 to 44.4%). At this point, the amino acids of CelB2 were replaced manually with the corresponding residues from EglA based on multiple sequence alignment results for family 12 enzymes (Fig. 1) using *PRONTO* (Laczkowski *et al.*, 1996).

Because of significant differences in loop regions between CelB2 and EglA, the core-tracing algorithm (Swanson, 1994) greatly assisted with the assignment of backbone location in initial approximately phased electron-density maps. The refinement proceeded with simulated annealing, which corrected errors introduced by the manual rebuilding and optimized the fit of the retraced region. After many steps of simulated-annealing refinement and manual rebuilding followed by refinement of individual *B* factors, the *R* value was reduced to 23.1% (R_{free} to 24.8%). Addition of water molecules to the model and final refinement reduced the *R* factor to 17.5% (R_{free} to 19.6%).

The structure of the palladium–EglA complex was determined using the data collected from a crystal of EglA soaked in 5 m*M* $PdCl_2$ for 20 min. The final model of the native EglA was used as an initial model for the refinement. After addition of palladium ions and water molecules along with several steps of simulated-annealing refinement, the final model with an *R* factor of 21.4% (R_{free} = 24.6%) was obtained (Table 1).

3. Results and discussion

3.1. Crystallography

The structure of EglA was determined by the Patterson search method using the structure of native endoglucanase from *S. lividans* (CelB2), with 26% identity with EglA, as a search model. The final structure contains one molecule per asymmetric unit and 16 asymmetric units per cell in the space group *I*422, with unit-cell parameters $a = b = 131.77$, $c = 71.72$ Å, $\alpha = \beta = \gamma = 90^\circ$ (Table 1). All 223 residues (1717 non-H protein atoms) and 124 water molecules are included in the refined model. The final crystallographic *R* value is 17.5% and the final R_{free} is 19.6% for the resolution range 30.0–2.1 Å. Using the program *PROCHECK* (Laskowski *et al.*, 1993), 90% of the non-glycine residues have their backbone conformational angles (φ/ψ) in the most favored regions of the Ramachandran plot (Ramakrishnan & Ramachandran, 1965). No residues fall into the disallowed regions.

A data set from a crystal soaked in 5 m*M* $PdCl_2$ was used to determine the structure of the palladium–EglA complex (Table 1). The final model consists of 223 residues, three Pd^{2+} ions and 34 water molecules. The structure was refined to a final *R* factor of 20.8% (R_{free} = 24.2%), including all reflections between 30.0 and 2.5 Å (98.5% completeness). The overall geometry is good (0.009 r.m.s.d. in bond lengths and 1.43° in bond angles). The root-mean-square fit between the Cα atoms of native EglA and the palladium–EglA complex is 0.4 Å.

3.2. Molecular structure

The catalytic core of EglA is a single-domain polypeptide chain with approximate dimensions of 40 × 40 × 35 Å. EglA has a 'jelly-roll' fold with two antiparallel β-sheets (outer *A* and inner *B*), which is a common folding motif in endoglucanases from family 12 (Fig. 2). β-Sheet *A* contains six strands (*A*1–*A*6) and β-sheet *B* contains nine strands (*B*1–*B*9). The hydrophilic face of β-sheet *A* is exposed to solvent. The hydrophilic face of β-sheet *B* forms a long open cleft which is the binding site for substrate.

One of the rims of the active-site cleft consists of two loops between strands *B*5 and *B*6 and between strands *B*7 and *B*8. Backbone atoms in the loop *B*7–*B*8 have higher main-chain temperature factors (40.3 $Å^2$ on average compared with 21.3 $Å^2$ for the average *B* factors of protein backbone atoms), suggesting that this loop is relatively flexible. This loop may move inward upon binding of the substrate and cover the binding site. Loops between strands *B*2 and *A*2 and between strands *B*3 and *A*3 form a part of the other rim of the active-site cleft.

There is a surface loop between β-strands *B*6 and *B*9. This loop, conserved in family 11 xylanases and family 12 endoglucanases, closes the substrate-binding cleft at its far end and has been referred to as the 'cord' (Torronen *et al.*, 1994). The residues in the cord region (Gly125, Asn126, Ile127, Gln128, Pro129, Ile130 and Gly131) have a clear electron-density map, indicating that the cord segment has a well defined structure. In family 12, amino acids in the cord region are highly conserved.

Two mostly conserved cysteine residues, Cys4 and Cys32, form a disulfide bridge to stabilize the three-dimensional fold of EglA by connecting strands *A*1 and *A*2. The structure contains two *cis*-prolines, Pro14 and Pro210. Pro14 is located in a turn between β-strands *B*1 and *B*2 and Pro210 is located between β-strands *B*4 and *A*4. EglA contains 19 negatively charged residues (aspartate and glutamate) and 11 positively charged residues (lysine, arginine and histidine). There are three salt bridges: between His46 and Asp36, between Arg123 and Glu145, and between Arg159 and Glu157. None of these residues is conserved in family 12.

3.3. The active-site cleft

The active site is located in the cleft on the surface of β-sheet *B* as confirmed by the Pd^{2+}-binding site. The extent of the cleft was estimated by calculating the solvent-accessible surface, giving a length of 35 Å and a depth of 9 Å. The cleft is 9 Å wide for most of its extended length. The smallest width is about 4 Å, between Trp22 and Gln153. This 35 Å long cleft is likely to bind around six glucopyranose units, which can be numbered from the non-reducing end (subsite −3) to the reducing end (subsite +3). The directionality of the cleft from the non-reducing end to the reducing end was determined based on comparison of this structure and the structure of the 2-fluorocellotryiosyl complex of CelB2 (Sulzenbacher *et al.*,

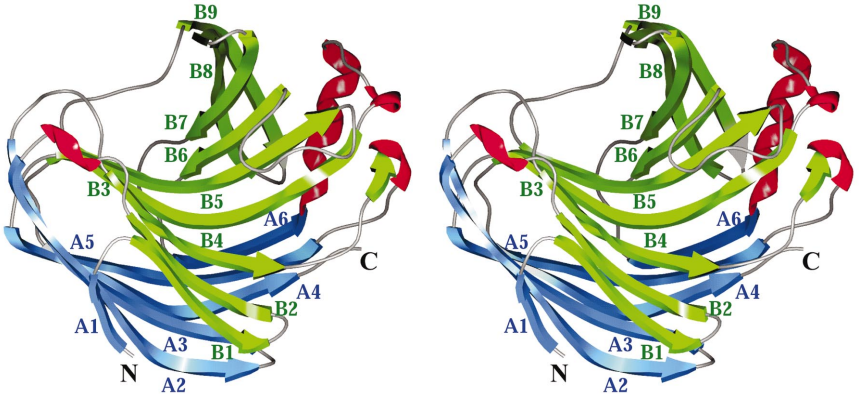

Figure 2
Ribbon stereo diagram of EglA. β-sheet *A* is in blue, β-sheet *B* is in green and helices are in red.

1999). The cleavage takes place, by definition, between subsites −1 and +1. An active site with a length of five or six glucopyranose units is consistent with the results of kinetics assays that showed the products of hydrolysis of CMC to be five or six glucopyranose units in length. The substrate-binding cleft is lined with a number of aromatic residues that may bind the hydrophobic faces of the pyranoside rings (Fig. 3).

By analyzing the conserved residues in the binding cleft and by comparison of the structure of EglA with the structure of a fluorocellotriosyl–CelB2 complex (Sulzenbacher *et al.*, 1999), we can make tentative functional assignments for the residues in the binding cleft (Fig. 3). At the non-reducing end of the cleft, the side chain of Tyr7 would interact with the glucopyranose ring in subsite −3. Trp22 is invariant and would flank subsite −2. Invariant residue Asn20 may interact with a hydroxyl group in the glucopyranose rings at either the −3 or −2 subsite. In subsite −1, the conserved aromatic residues Tyr61 and Phe101 may stabilize the hydrophobic glucopyranose ring. In subsite −1, stabilization would occur through hydrogen bonding to the acid/base Glu204 and to residues Glu158 and Trp22. The invariant Met118, sitting on the bottom of the binding cleft, and Phe206 are probably involved in a stacking interaction with a glycosyl moiety in the +1 subsite (Sulzenbacher *et al.*, 1999). The invariant Pro129 is tightly sandwiched between the aromatic rings of Trp120 and Trp147 in the middle of the 'cord' region. It has been suggested that the cord loop may undergo conformational changes upon substrate binding, with conformational flexibility accompanied by a prolyl *cis–trans* isomerization (Torronen *et al.*, 1994; Sulzenbacher *et al.*, 1999). It is presumed that the cord movement upon ligand binding facilitates the interaction of the conserved Trp120 and Trp147 with the hydrophobic faces of the oligosaccharide ligand in subsites +2 and +3, respectively.

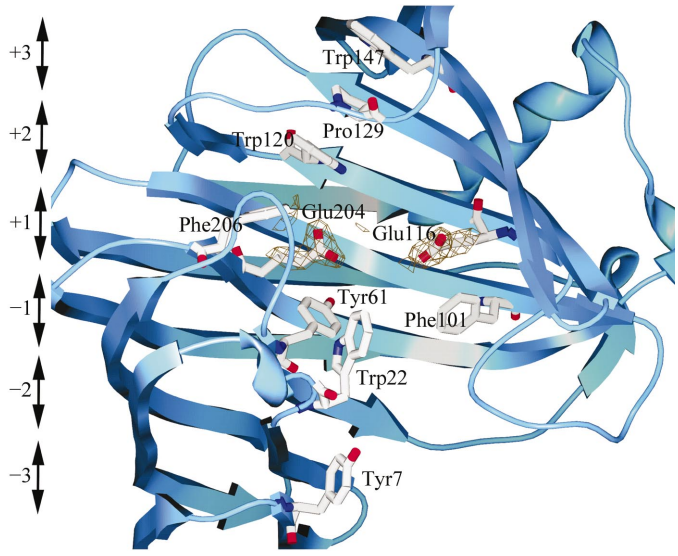

Figure 3
The binding-cleft environment of EglA. The electron-density map ($2F_o - F_c$) for the catalytic residues Glu116 and Glu204 is contoured at 2.0σ. The six putative subsites (from subsite −3 through +3) are shown by vertical arrows.

Family 12 enzymes perform catalysis with net retention of anomeric configuration. The mechanism is widely held to be a double-displacement reaction in which a covalent glycosyl-enzyme intermediate is formed and subsequently hydrolyzed *via* an oxocarbonium-ion-like transition state (Koshland, 1953; Davies *et al.*, 1998). Enzymological (Zechel *et al.*, 1998) and structural (Sulzenbacher *et al.*, 1999) studies of CelB2 confirmed the role of Glu120 and Glu203 as the catalytic nucleophile and the general acid/base, respectively. Glu120 and Glu203 in CelB2 correspond to Glu116 and Glu204 in EglA, respectively.

Asp99, which is conserved in family 12, is 2.6 Å from the nucleophile Glu112. This residue would not only help to maintain an appropriate charge on the nucleophile during the catalytic cycle, but would also hold Glu116 in a favorable orientation for nucleophilic attack. Asp95 is 2.6 Å from the catalytic acid/base residue Glu204. This residue, which is conserved or substituted by aspargine in other family 12 enzymes, may have a role in appropriate charge maintenance of catalytic acid/base residues during the enzymatic reaction.

3.4. Structural comparison of EglA with CelB2 and xylanase

The sequence identity between CelB2 (Sulzenbacher *et al.*, 1997) and EglA is relatively low at 26% (Fig. 1). Superposition of the refined structure of EglA and CelB2 (PDB code 1nlr) using the *SPOCK* program package (Christopher, 1998) resulted in an r.m.s.d. of 1.17 Å for 159 equivalent Cα positions limited to core regions. This three-dimensional comparison shows that the core regions (β-sheets *A* and *B*) are similar, whereas significant differences are observed in the surface-loop regions (Fig. 4*a*). The long loop between strands *B*3 and *A*5 in CelB2 contains both a 3_{10}-helix and two additional short β-strands (β-sheet *C*). These two short β-strands, together with the two loops between strands *B*5 and *B*6 and strands *B*7 and *B*8, make up one rim of the active-site cleft in CelB2. The residues His65 and Tyr66 in β-sheet *C* interact with the glucopyranose ring in subsites −2 and −3 (Sulzenbacher *et al.*, 1999). However, these two short β-strands do not exist in the EglA structure and thus β-sheet *C* of CelB2 is not conserved in all members of family 12. During preparation of this manuscript, the 1.9 Å resolution structure of a β-1,4-endoglucanase from *Trichoderma reesei* (Cel12A) was determined (Sandgren *et al.*, 2001) and is similar to the native structure described here. Cel12A also lacks β-sheet *C*. Since EglA does not possess β-sheet *C*, the active-site cleft is also open at the non-reducing end. Asn63 from β-strand *B*3 in EglA, corresponding to Tyr66 in CelB2, would interact with substrate in the subsite −3.

Although the sequence similarity between families 11 and 12 is extremely low, hydrophobic cluster analysis indicates that the structures would have the same overall fold (Torronen *et al.*, 1993). On the basis of sequence analysis, families 11 and 12 were proposed to be a single clan, GH-C. Superposition of EglA and xylanase II, XYNII, from *T. reesei* (PDB entry 1enx)

indicated an r.m.s.d. of 2.36 Å for 122 equivalent Cα positions in the core region of two proteins (Fig. 4*b*). There is 16% identity between the sequences of EglA and XYNII. Significant differences are found at the surface-loop regions. While XYNII lacks the putative β-strand *A*1, there is a short *A*1 strand in EglA. XYNII, like EglA, lacks the two short β-strands between strands *B*3 and *A*5 (β-sheet *C*). Therefore, the non-reducing end of the binding cleft of XYNII, like EglA and Cel12A, is open.

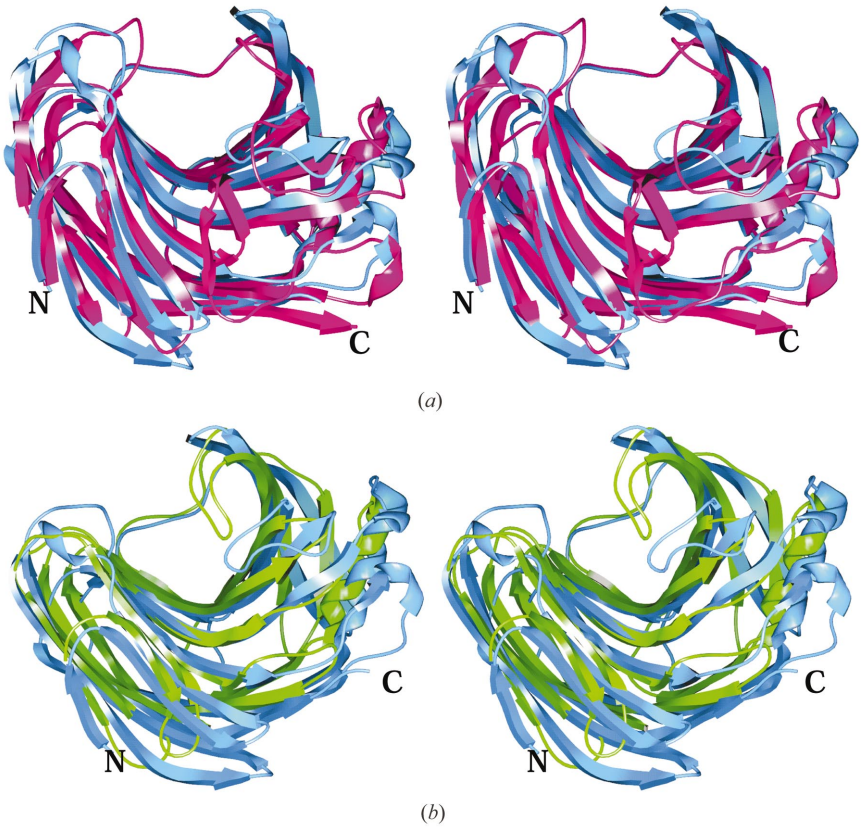

Figure 4
Stereoview of the superposition of EglA with (*a*) CelB2 from *S. lividans* and (*b*) XYNII from *T. reesei*. EglA is in blue, CelB2 is in purple and XYNII is in green.

3.5. Structure of palladium–EglA complex

Addition of 100 μ*M* $PdCl_2$ to 1 μg ml^{-1} EglA inhibits more than 94% of enzyme activity against 1% carboxymethyl cellulose (CMC) (Fig. 5). The inhibition of EglA by palladium is irreversible based on the following findings. When 100 μ*M* $PdCl_2$ is incubated with 5 m*M* EDTA for 15 min before addition of EglA, no inhibition of EglA will occur. This shows that EDTA chelates with Pd^{2+} and subsequently prevents Pd^{2+} from inhibiting EglA. However, if 5 m*M* EDTA is added 10 min after the start of reaction between EglA and CMC in the presence of 100 μ*M* $PdCl_2$, there is no reversal of inhibition even after 60 min of reaction time. Apparently, in this condition Pd^{2+} binds irreversibly to EglA and addition of EDTA does not remove Pd^{2+} from its binding site. EDTA by itself does not have any effect on the activity of the EglA. Furthermore, platinum, mercury or lead complexes did not show an inhibitory effect on the enzyme.

The structure of the palladium–EglA complex shows that three Pd^{2+} ions bind to each molecule of EglA (Fig. 6*a*). Superposition of the Cα atoms of the native EglA and palladium–EglA complexes shows an r.m.s.d. of 0.4 Å. Therefore, inactivation of EglA by palladium does not arise from conformational changes in the backbone of the enzyme.

A Pd^{2+} ion binds to the active site of EglA by forming a coordinate covalent bond with Met118 S^{δ} (Fig. 6*b*). The bond length is almost 2.2 Å, which is consistent with bond lengths between Pd^{2+} and thioether atoms in other organometallic compounds (Battaglia *et al.*, 1973; McCrindle *et al.*, 1982). Met118, which is conserved in all members of family 12, resides on the bottom of the binding cleft and its side chain is located between the general acid/base (Glu204) and the catalytic nucleophile residue (Glu116) involved in a stacking interaction with a glycosyl moiety in the +1 subsite (Sulzenbacher *et al.*, 1999). This Pd^{2+} ion forms another coordinate covalent bond with the nucleophilic Glu116 $O^{\varepsilon 1}$ (∼2.4 Å bond length). It seems that Met118 is necessary for the binding of Pd^{2+} to EglA. However, for inactivation of the enzyme, binding to the nucleophilic glutamate is important. The

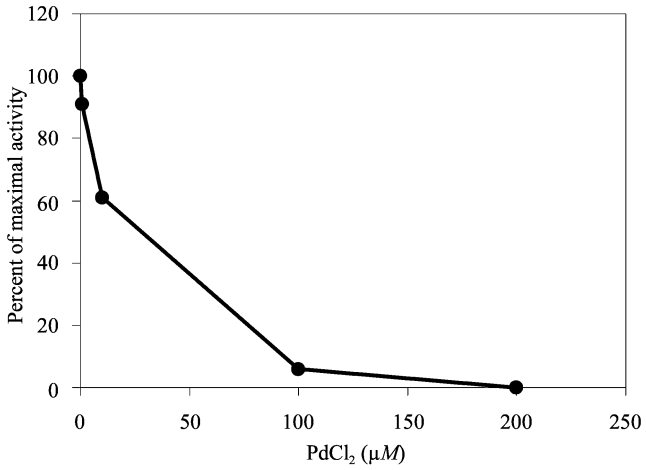

Figure 5
The inhibitory effect of $PdCl_2$ on EglA. Solutions of 1 μg ml^{-1} EglA (1 μg ml^{-1}) were incubated with 0, 1, 10, 100 and 200 μ*M* $PdCl_2$ for 15 min at room temperature. The enzyme activity was determined against 1% carboxymethyl cellulose (CMC).

complex between Pd^{2+} and EglA is a σ-bonded organopalladium (II) complex, as found in many organopalladium complexes. We could not detect any other ligand involved in coordination. This finding is consistent with previous assumption that $PdCl_2$ binds to the endoglucanase as a free

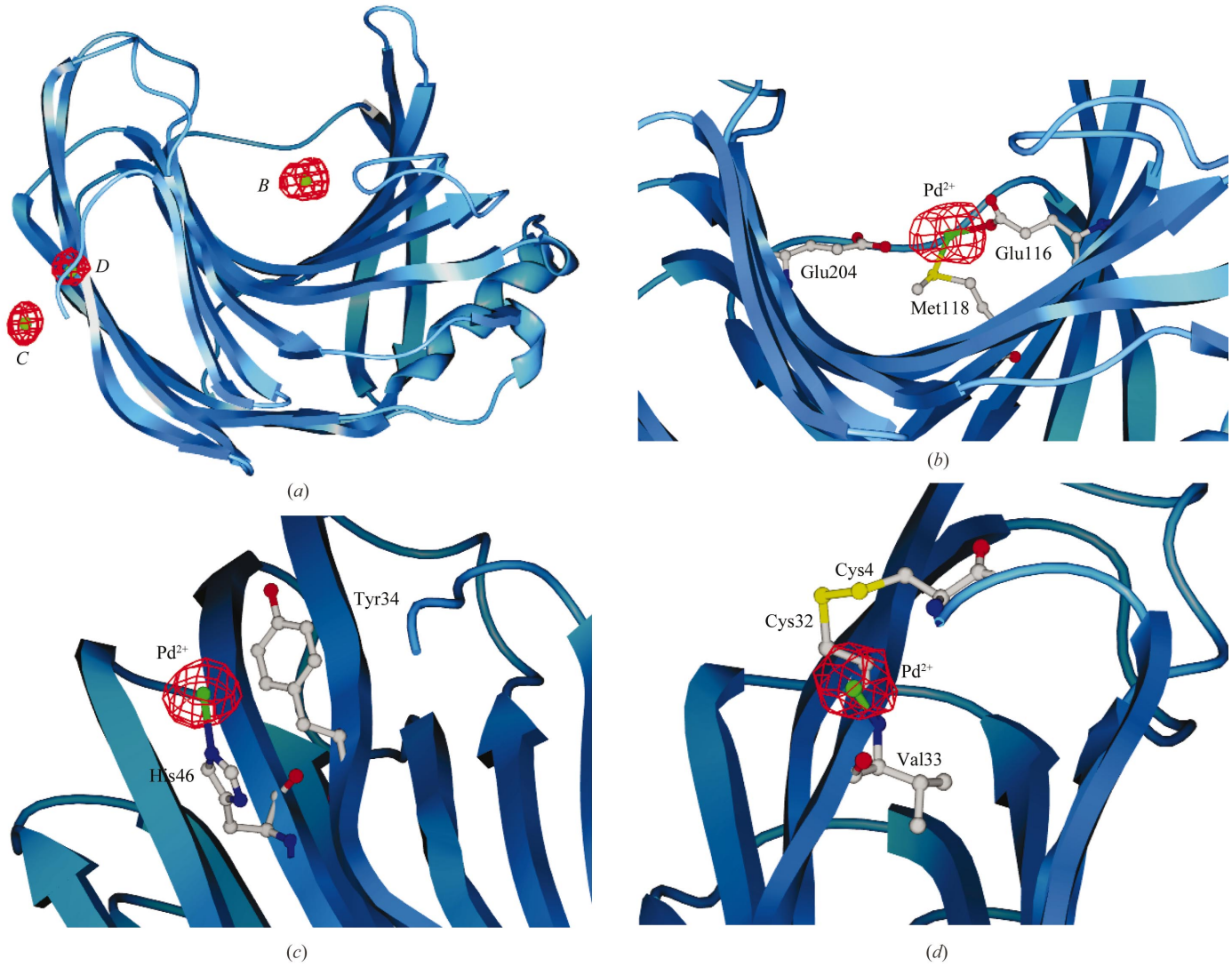

Figure 6
(*a*) The structure of EglA complexed with three Pd^{2+} ions, *B*, *C* and *D*. (*b*), (*c*) and (*d*) The binding site for Pd^{2+} ions *B*, *C* and *D*, respectively. The electron-density maps ($F_o - F_c$) for the three Pd^{2+} ions are contoured at 7.0σ. The Pd^{2+} ions are in green.

Pd^{2+} ion (Lassig *et al.*, 1995). In addition to those coordinate covalent bonds, Pd^{2+} makes non-bonded interactions with Glu116 $O^{\varepsilon 2}$ (~2.9 Å) and with the catalytic acid/base residue Glu204 $O^{\varepsilon 1}$ (~3.2 Å). Based on these results, one can conclude that the coordinate covalent binding of Pd^{2+} to Met118 S^{δ} and Glu116 $O^{\varepsilon 1}$ blocks the active site and inactivates the enzyme.

The second Pd^{2+} ion forms a coordinate covalent bond with His46 $N^{\varepsilon 2}$ on the surface of the protein (Fig. 6*c*). Since His46 is far from the active site of the enzyme, binding of palladium to this site does not have any inhibitory effect on the enzyme. These results are consistent with the previous kinetics results showing that palladium most likely forms a covalent bond with the imidazole N atom of His206 or His228 of CBH I, but that this binding does not play any significant role in the inhibition of CBH I by palladium (Lassig *et al.*, 1995). This palladium ion makes an electrostatic interaction with the phenyl ring π-electrons of Tyr34, which is about 2.6 Å from the Pd^{2+} ion.

The third Pd^{2+} ion forms coordinate covalent bonds with Val33 N (Fig. 6*d*). This palladium ion is close to the only disulfide bond in EglA between Cys4 and Cys32 (the distance between the Pd^{2+} ion and Cys4 S^{δ} is ~3.3 Å). Lassig *et al.* (1995) proposed that inactivation of CBH I by palladium is a consequence of the cleavage of disulfide bonds. However, our structural studies of the EglA–palladium complex show that the disulfide bond between Cys4 and Cys32 has not been cleaved, although the palladium-binding site is close to this disulfide bond.

The results of this study establish the mode of inhibition of EglA by palladium complexes. However, inhibition of other cellulases by palladium complexes may follow alternative mechanisms. Cellobiohydrolase I (CBH I) and endoglucanase II (EG II) from *T. reesei*, which are also inhibited by palladium complexes, belong to family 7 and family 5, respectively. These two enzymes do not have methionine in their active sites; therefore, the mechanism of inhibition in these two enzymes

should be different from that of EglA. Structural elucidation of other cellulase–palladium complexes will show the mechanism of their inactivation by palladium complexes.

The structure of the palladium–EglA complex now adds a new dimension to efforts aimed towards designing novel inhibitors in order to control cellulase activity in the fungus *A. niger* which deteriorates cellulose-based materials (Reese, 1974).

We thank Professor Dr Dimiter Kolev for his persuasive encouragement to pursue this project 12 years ago. We thank Professor Jay Fox for his generous willingness to supply us with partial sequences and Dr Mike Himmel for partial financial support. Significant financial support was provided by the Robert A. Welch Foundation (A-328). We thank Dr Istvan Botos for technical assistance. We thank Professor Dr Gerd Folkers for his exceptional willingness to assist us with personal aspects of our research efforts and dedicate this paper to him.

References

Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
Bakalova, N., Petrova, S. & Kolev, D. (1996). *Pharmazie*, **51**, 761–764.
Battaglia, L. P., Corradi, A. B., Palmieri, C. G., Nardelli, M. & Tani, M. E. V. (1973). *Acta Cryst.* B**29**, 762–767.
Bhat, K. M. & Bhat, S. (1997). *Biotechnol. Adv.* **15**, 583–620.
Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
Brünger, A. T. (1997). *Methods Enzymol.* **277**, 558–580.
Christopher, J. A. (1998). *SPOCK: The Structural Properties Observation and Calculation Kit.* The Center for Macromolecular Design, Texas A&M University, College Station, TX, USA.
Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schulein, M. & Whiters, S. G. (1998). *Biochemistry*, **37**, 11707–11713.
DeLano, W. L. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 740–748.
Eriksson, K. E. & Pettersson, B. (1975). *Eur. J. Biochem.* **51**, 193–206.
Gubitz, G. M., Mansfield, S. D., Bohm, D. & Saddler, J. N. (1998). *J. Biotechnol.* **65**, 209–215.
Henrissat, B. & Bairoch, A. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.
Jue, C. K. & Lipke, P. N. (1985). *J. Biochem. Biophys. Methods*, **11**, 109–115.
Koshland, D. E. J. (1953). *Biol. Rev.* **28**, 416–436.
Laczkowski, A., Meyer, E. F., Swanson, S. M. & Worick, R. K. (1996). *J. Iran Univ. Med. Sci.* **3**, 31–41.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lassig, J. P., Shultz, M. D., Gooch, M. G., Evans, B. R. & Woodward, J. (1995). *Arch. Biochem. Biophys.* **322**, 119–126.
Lynd, L. R., Wyman, C. E. & Gerngross, T. U. (1999). *Biotechnol. Prog.* **15**, 777–793.
McCrindle, R., Ferguson, G., McAlees, A. J., Parvez, M. & Stephenson, D. K. (1982). *J. Chem. Soc. Dalton Trans.* **7**, 1291–1296.
Navaza, J. & Vernoslova, E. (1995). *Acta Cryst.* A**51**, 445–449.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* A**52**, 659–668.
Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
Reese, E. T. (1974). *Biotechnol. Bioeng. Symp.* **6**, 9–20.
Sandgren, M., Shaw, A., Ropp, T. H., Wu, S., Bott, R., Cameron, A. D., Stahlberg, J., Mitchinson, C. & Jones, T. A. (2001). *J. Mol. Biol.* **308**, 295–310.
Schulein, M., Kauppinen, M. S., Laneg, L., Lassen, S. F., Andersen, L. N., Klysner, S. & Nielson, J. B. (1998). *ACS Symp. Ser.* **687**, 66–74.
Sheehan, J. & Himmel, M. (1999). *Biotechnol. Prog.* **15**, 817–827.
Shultz, M. D., Lassig, J. P., Gooch, M. G., Evans, B. R. & Woodward, J. (1995). *Biochem. Biophys. Res. Commun.* **209**, 1046–1052.
Sulzenbacher, G., Mackenzie, L. F., Wilson, K. S., Wither, S. G., Dupont, C. & Davies, G. J. (1999). *Biochemistry*, **38**, 4826–4833.
Sulzenbacher, G., Shareck, F., Morosoli, R., Dupont, C. & Davies, G. J. (1997). *Biochemistry*, **36**, 16032–16039.
Swanson, S. (1994). *Acta Cryst.* D**50**, 695–708.
Torronen, A., Harkki, A. & Rouvinen, J. (1994). *EMBO J.* **13**, 2493–2501.
Torronen, A, Kubicek, C. P. & Henrissat, B. (1993). *FEBS Lett.* **321**, 135–9.
Wyman, C. E. (1999). *Annu. Rev. Energy Environ.* **24**, 189–226.
Zechel, D. L., Shouming, H. E., Dupont, C. & Withers, S. G. (1998). *Biochem. J.* **336**, 139–145.