Engineering of a glycosidase Family 7 cellobiohydrolase to more alkaline pH optimum: the pH behaviour of *Trichoderma reesei* Cel7A and its E223S/A224H/L225V/T226A/D262G mutant

Dieter BECKER*, Christophe BRAET†, Harry BRUMER III‡, Marc CLAEYSSENS†, Christina DIVNE‖, B. Richard FAGERSTRÖM§, Mark HARRIS‖, T. Alwyn JONES‖, Gerard J. KLEYWEGT‖, Anu KOIVULA¶, Sabah MAHDI**, Kathleen PIENS†, Michael L. SINNOTT*[1], Jerry STÅHLBERG**, Tuula T. TEERI‡, Melanie UNDERWOOD* and Gerd WOHLFAHRT¶

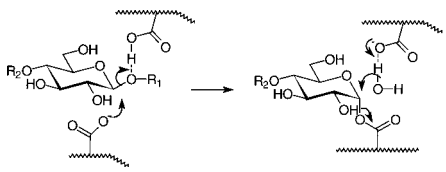
*Department of Paper Science, University of Manchester Institute of Science and Technology, P.O. Box 88, Sackville Street, Manchester M60 lQD, U.K., †Department of Biochemistry, Physiology and Microbiology, University of Ghent, Ledeganckstraat 35, B-9000 Ghent, Belgium, ‡Department of Biotechnology, Royal Institute of Technology, S-10044 Stockholm, Sweden, §Röhm Enzyme Finland OY, PL 26, Tykkimäentie 15, FIN-05200 Rajamäki, Finland, ‖Department of Cell and Molecular Biology, Uppsala University, BMC, P.O. Box 596, SE-75124, Uppsala, Sweden, ¶VTT Biotechnology, P.O. Box 1500, FIN-02044 VTT, Espoo, Finland, and **Department of Molecular Biology, Swedish University of Agricultural Sciences, BMC, P.O. Box 590, SE-75124, Uppsala, Sweden

The crystal structures of Family 7 glycohydrolases suggest that a histidine residue near the acid/base catalyst could account for the higher pH optimum of the *Humicola insolens* endoglucanase Cel7B, than the corresponding *Trichoderma reesei* enzymes. Modelling studies indicated that introduction of histidine at the homologous position in *T. reesei* Cel7A (Ala^{224}) required additional changes to accommodate the bulkier histidine side chain. X-ray crystallography of the catalytic domain of the E223S/A224H/L225V/T226A/D262G mutant reveals that major differences from the wild-type are confined to the mutations themselves. The introduced histidine residue is in plane with its counterpart in *H. insolens* Cel7B, but is 1.0 Å (= 0.1 nm) closer to the acid/base Glu^{217} residue, with a 3.1 Å contact between $N^{\epsilon 2}$ and $O^{\epsilon 1}$. The pH variation of k_{cat}/K_m for 3,4-dinitrophenyl lactoside hydrolysis was accurately bell-shaped for both wild-type and mutant, with pK_1 shifting from 2.22 ± 0.03 in the wild-type to 3.19 ± 0.03 in the mutant, and pK_2 shifting from 5.99 ± 0.02 to 6.78 ± 0.02. With this poor substrate, the ionizations probably represent those of the free enzyme. The relative k_{cat} for 2-chloro-4-nitrophenyl lactoside showed similar behaviour. The shift in the mutant pH optimum was associated with lower k_{cat}/K_m values for both lactosides and cellobiosides, and a marginally lower stability. However, k_{cat} values for cellobiosides are higher for the mutant. This we attribute to reduced non-productive binding in the +1 and +2 subsites; inhibition by cellobiose is certainly relieved in the mutant. The weaker binding of cellobiose is due to the loss of two water-mediated hydrogen bonds.

Key words: cellulase, cellulose, endoglucanase, enzyme kinetics, pH-dependence.

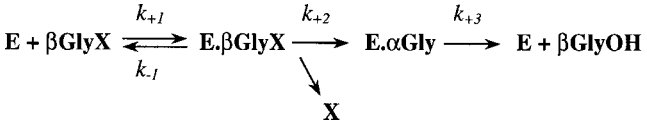
INTRODUCTION

The catalytic mechanism of retaining glycosidases, that is, those glycoside hydrolases which give the product reducing sugar in the same anomeric configuration as the starting glycoside, is now well understood. The reaction is a double displacement, involving a covalent intermediate of anomeric configuration opposite to that of the substrate, with both chemical steps involving oxo-carbenium-ion-like transition states [1]. In the commonest type of such enzyme, the nucleophile is a carboxylate group (Asp or Glu) of the protein, whereas a second carboxylic acid group partially protonates the leaving-group oxygen atom in the first step, and partially deprotonates the incoming water molecule in the second step (Scheme 1):



Scheme 1 Commonest chemical mechanism of a retaining β-glycosidase

The requirement for two carboxylate groups, one protonated and the second deprotonated, in the first chemical step of the enzyme, predicts a classical bell-shaped pH profile for that step. The acid limb of the bell is commonly attributed to protonation of the nucleophile, and the alkaline limb to the deprotonation of the acid/base catalyst. For many enzymes the pH behaviour of the steady-state parameter k_{cat}/K_m is equivalent to, and far more accessible than, the pH behaviour of the individual rate constants. For a retaining glycosidase acting on a glycosyl derivative βGly-X the kinetic scheme becomes that of Scheme 2:



Scheme 2

which gives the steady-state kinetic parameters given by eqns 1–3:

Abbreviations used: BMCC, bacterial microcrystalline cellulose; CBH, cellobiohydrolase; CNP, 2-chloro-4-nitrophenyl; DNP, 3,4-dinitrophenyl; EG, endoglucanase; ES, enzyme–substrate; *H. i.*, *Humicola insolens*; MU-Lac, 4-methylumbelliferyl β-D-lactoside; RMS, root mean square; *T. r.*, *Trichoderma reesei*; wt, wild-type; E223S etc., $Glu^{223} \rightarrow Ser$ etc.; PDB, Protein Data Bank.

[1] To whom correspondence should be addressed (e-mail: Michael.Sinnott@umist.ac.uk)

$$k_{cat} = \frac{k_{+2} \cdot k_{+3}}{k_{+2} + k_{+3}} \quad (1)$$

$$K_m = \frac{k_{-1} + k_{+2}}{k_{+1}} \cdot \frac{k_{+3}}{k_{+2} + k_{+3}} \quad (2)$$

$$\frac{k_{cat}}{K_m} = \frac{k_{+1} \cdot k_{+2}}{k_{-1} + k_{+2}} \quad (3)$$

In the case of a slow substrate which is 'non-sticky' (i.e. $k_{-1} \gg k_{+2}$), then the pH variation of k_{cat}/K_m is that of k_{+2} and K_s. If the 'bottleneck' assumption of the classic treatment of Alberty and Bloomfield [2] holds, that at some point in the mechanism only one protonation-state of complex continues to product, then the pH variation of k_{cat}/K_m is affected only by the ionizations of free enzyme and of free substrate. In contrast, the pH variation of the separated parameters k_{cat} and K_m will be far more difficult to interpret, as these parameters are blends of rate constants, each one of which will have a separate pH-dependence.

The hydrolysis of cellulose presents evolution with the physical problem of attack on an insoluble substrate, as well as the chemical problem of cleavage of a glycosidic linkage. The solution to the combined physical and chemical problem requires cellulolytic organisms to produce an array of various $\beta(1 \rightarrow 4)$glucan-hydrolysing activities. With some cellulolytic bacteria the different enzyme activities are associated in a large multi-enzyme complex known as a cellulosome [3], whereas with other bacteria and all aerobic cellulolytic fungi the enzymes are not associated.

The best-studied cellulase system is from the filamentous soft-rot fungus *Trichoderma reesei* (*T. r.*). The fungus produces at least seven cellulase components, two cellobiohydrolases and a range of endoglucanases. The gross molecular architecture of all the cellulase components is similar (except the small endoglucanase Cel12A, formerly called 'EG III' [4]), a catalytic domain that is attached through a heavily *O*-glycosylated linker region to a cellulose-binding domain [5]. These enzymes, whether as mixtures or as individual components, are probably the most widely used cellulases industrially, with applications in the textile industry and in the pulp and paper industry for de-inking and refining. For most of these applications, however, the pH optima (≈ 5) are too acid: for pulp applications the enzymes should ideally be optimally active at neutral-to-alkaline pH and for textile (detergent) applications at alkaline pH.

The present paper reports the successful protein engineering of the most abundant component of the *T. r.* cellulase complex, cellobiohydrolase Cel7A, to more alkaline pH optimum. *T. r.* Cel7A {formerly called cellobiohydrolase (CBH) I [4]} is placed in Family 7 of the Henrissat and Bairoch [6] classification of glycosidase sequences. This sequence family has a perfect correlation with the stereochemistry of the catalysed reaction (in this case retention), and an excellent correlation with overall protein fold [7]. Family 7 contains both cellobiohydrolases and endoglucanases, the structures of the cellobiohydrolases differing from the endoglucanases by the existence of loops of polypeptide chain covering the active-site residues and converting the active-site cleft of the endoglucanases into the characteristic tunnel of the cellobiohydrolases [8]. This topology is almost certainly related to the crystal-lattice-disrupting action of the cellobiohydrolases; a recent X-ray study of complexes of *T. r.* Cel7A with a range of ligands has revealed no less than 10 saccharide-binding sites [9], which define a pathway for the glucan chain as it threads its way through the catalytic domain. The tunnel apparently promotes processive action, i.e. the repeated removal of consecutive cellobiose units from the end of a cellulose chain. Interestingly, the two cellobiohydrolases in *T. r.* seem to work in opposite directions along cellulose chains, Cel7A from the reducing end and Cel6A (formerly CBH II) from the non-reducing end [8,10,11].

Table 1 Numbering of corresponding amino acid residues and sequence differences at the active site in cellobiohydrolase Cel7A and endoglucanase Cel7B from *T. reesei* and endoglucanase Cel7B from *H. insolens*

		Sequence						
	T. r. Cel7A numbering ...	223	224	225	226	227	228	262
Enzyme	*H. i.* Cel7B numbering ...	208	209	210	211	212	213	240
T. r. Cel7A wt		Glu	Ala	Leu	Thr	Pro	His	Asp
T. r. Cel7B		Asn	Ala	Leu	Thr	Pro	His	Gly
H. i. Cel7B		Ser	His	Val	Ala	Pro	His	Gly
T. r. Cel7A mutant		Ser	His	Val	Ala	Pro	His	Gly

In addition to Cel7A, the *T. r.* cellulase complex contains another Family-7 enzyme, namely endoglucanase Cel7B (formerly EG I [4]), whose structure has been solved [12]. Comparison of the structure with that of another Family-7 endoglucanase, Cel7B of *Humicola insolens* (*H. i.*) [13], which has a broad pH optimum (activity invariant between pH 5.5 and 7.5) and useful activity even at pH 9 [14], revealed that a key difference in the structure was the existence of a histidine residue within 4.3 Å of the acid/base carboxylate group, which was not present in *T. r.* Cel7A or *T. r.* Cel7B. Hydrogen-bonding between the catalytic acid/base and the histidine residue was considered a possibility, at least through a water molecule. It was hypothesized [12] that this histidine residue at position 209 in *H. i.* Cel7B (Cel7B numbering) could act as a proton source for the catalytic acid/base Glu^{202} at high pH. Both of the *T. r.* enzymes have an alanine residue at this position.

To test the hypothesis, a mutant of *T. r.* Cel7A was produced that had a histidine residue at the equivalent position (224 in the Cel7A numbering scheme). In addition to the $Ala^{224} \rightarrow His$ mutation, four adjacent residues were replaced by their *H. i.* Cel7B counterparts. Thr^{226} was changed to Ala and Asp^{262} to Gly, in order to accommodate the increased bulk of the histidine. The other two residues, on either side next to the histidine in the sequence, point in the opposite direction on the β-strand, into a small cavity in the β-sheet interface. They were replaced by their *H. i.* Cel7B counterparts in order to allow for possible main-chain adjustments at the histidine. Glu^{223} was replaced with Ser and Leu^{225} with Val. The changed residues of the site-directed *T. r.* Cel7A mutant and the corresponding residues of the wild-type (wt) enzymes discussed in its design are set out in Table 1.

EXPERIMENTAL

Construction of the mutant clone and transformation to *T. reesei*

Escherichia coli strain DH5α (Promega) was used as the cloning host for the DNA constructs and a *T. r.* strain lacking the genes for Cel7A and Cel7B as the production host of the mutant. The pEM-F5 plasmid, containing the *cel7A* gene under its own promoter [15], was used as an expression host vector. For selection of *T. r.* transformants, hygromycin selection plasmid pRLMex30 was used [16]. The construction of the mutant expression plasmid was done by PCR overlap extension method,

and the DNA sequence of the whole mutated area was subjected to DNA sequencing. *T. r.* transformation and choosing the best producing transformant was performed basically as described by Ståhlberg et al. [17].

Isolation of wt and mutant enzymes

wt Cel7A from *T. r.* strain QM 9419 was from the cultivation and purification described previously [17]. The Cel7A wt used for kinetic measurements was further purified by affinity chromatography using *p*-aminobenzyl 1-thio-β-cellobioside as ligand coupled to Sepharose 4B, followed by anion-exchange chromatography as described in [18].

The purification of the *T. r.* Cel7A mutant (E223S, A224H, L225V, T226A and D262G) was started by adjusting the culture filtrate to pH 5.2 with 0.1 M NaOH and diluting to 1.9 mS/cm with water. The sample was then applied to a DEAE-Sepharose FF column (5 cm × 18 cm) equilibrated with 20 mM sodium acetate, pH 5.2. The column was eluted with a linear gradient of 0 to 0.5 M NaCl in equilibration buffer. The fractions containing the highest activity against 4-methylumbelliferyl β-D-lactoside (MU-Lac) were pooled and adjusted to pH 6 with 1 M potassium phosphate buffer, pH 6.0. $(NH_4)_2SO_4$ was added to 0.5 M and the sample was loaded on a Phenyl-Sepharose FF column (5 cm × 10 cm) equilibrated with 20 mM potassium phosphate (pH 6.0)/0.5 M $(NH_4)_2SO_4$. Elution was performed with a linear gradient to 20 mM potassium phosphate, pH 6.0. Analysis of the fractions was performed as described above, using MU-Lac as a substrate as well as subjecting material to SDS/PAGE and Western blot (see below). The purest fractions, eluted after extensive washing with 20 mM potassium phosphate after the gradient, were combined. This pool was concentrated by $(NH_4)_2SO_4$ (45%, w/v) precipitation. The precipitate was suspended in 50 mM sodium acetate buffer, pH 5.0, and applied to a Superdex G-75 HiLoad column equilibrated with the same buffer. The fractions containing the highest protein concentration were pooled and used in the present study.

In both cases the purified mutant or wt enzyme was subjected to Western-blot analysis and detected with a monoclonal antibody against *T. r.* Cel7A. The contaminating activities on a small chromophoric substrate (4-MU β-D-glucoside) and on hydroxyethylcellulose were also checked as described previously [19]. Protein concentration for both Cel7A preparations was measured by UV absorption at 280 nm using $\epsilon_{280} = 83000\ M^{-1} \cdot cm^{-1}$ determined by quantitative amino acid analysis (A. Koivula, unpublished work).

The catalytic domain of the Cel7A mutant, used for X-ray crystallography, was prepared by proteolytic cleavage of the intact enzyme with papain. The first attempt, using the previously established protocol for domain cleavage of Cel7A {overnight incubation at room temperature in 50–100 mM sodium acetate, pH 5.0, with 100:1 (w/w) ratio of Cel7A to papain [17,18,20]} resulted in the complete degradation of the enzyme. After trying shorter incubations with small aliquots (20 min, 1, 2, and 4 h) the preparative cleavage was performed as follows: papain was activated by adding 10 μl of a crystalline suspension (10 mg/ml; Boehringer-Mannheim) to 1 ml of activation buffer [0.1 M sodium phosphate (pH 7.0)/2 mM EDTA/2 mM dithioerythritol] and incubated 30 min at room temperature. The papain solution was mixed with 13 mg of intact Cel7A mutant protein in 7 ml of 0.1 M sodium acetate, pH 5.0 (8 ml reaction mixture). After 75 min at room temperature (approx. 22 °C) the papain was inactivated by the addition of 10 μl of 0.5 M iodoacetic acid in water. The digest was fractionated by anion-exchange chromatography (Pharmacia Source 30Q; 1.6 cm × 10 cm; linear gradient of 0–0.5 M NaCl in 20 mM sodium acetate, pH 4.0). The major peak, containing only the catalytic domain of Cel7A (as judged by SDS/PAGE) was diafiltered to 10 mM sodium acetate, pH 5.0, concentrated to 1.2 mg protein/ml and stored frozen at −20 °C.

Table 2 Crystallographic data and structure refinement statistics

Parameter	Value
Crystallographic data*	
Space group:	I222, one molecule per asymmetric unit
Cell axes [*a*, *b*, *c* (Å); α, β, γ]	83.1, 83.3, 110.6; 90°, 90°, 90°
Resolution limits (Å)	30.0–1.6
Resolution limits, outer shell (Å)	1.63–1.6
No. of unique reflections	50924
Average multiplicity	4.7 (4.5)
Completeness (%)	100 (99.9)
R_{merge} (%)	6.4 (23.7)
$\langle I/\sigma\ (I)\rangle$	24.6 (5.5)
Structure refinement	
Refinement resolution range (Å)	20.0–1.6
No. of non-hydrogen atoms	3596
R-value, R_{free}	24.0, 26.6
Mean B-value (Å^2)	
Protein atoms	7.4
Solvent atoms	14.6
Glucosyl atoms	11.7
Ramachandran outliers (%)†	0

* Statistics for highest-resolution shell in parenthesis.
† Percentage of residues that fall outside core regions of the Ramachandran plot.

X-ray crystallography: crystallization and data collection

Crystallization was performed at room temperature via hanging-drop vapour diffusion [21]. A 3 μl portion of protein solution containing 1.2 mg/ml of the catalytic domain of Cel7A mutant in 10 mM sodium acetate, pH 5.0, was mixed with an equal volume of a reservoir solution containing 18.75% poly(ethylene glycol) 5000 monomethyl ether (Fluka), 12.5% glycerol, 20 mM sodium Mes buffer, pH 6.0, and 20 mM $CoCl_2$. The drops were left to equilibrate against 1 ml of the reservoir for 2 days before microseeding was performed using previously prepared crystals of the same protein. Crystals started to appear within 5 hours of seeding and were left to grow for another 2 weeks before use. The crystal used for data collection was approx. 0.3 mm × 0.2 mm × 0.1 mm and was extracted from the crystallization drop with a 0.5 mm-diameter cryo-loop (Hampton Research, Laguna Niguel, CA, U.S.A.). It was then mounted directly on the goniometer of our in-house Rigaku/R-AXIS diffractometer equipped with an Oxford system cryo-cooling device, which was used to flash-freeze the crystal. After checking the diffraction, the crystal was stored under liquid nitrogen for transportation to the synchrotron.

A single X-ray dataset was collected on beam line X11 of the Deutsches Elektronen Synchrotron ('DESY') synchrotron at the European Molecular Biology Laboratory ('EMBL') outstation in Hamburg, Germany ($\lambda = 0.9050$ Å), using an MAR Research Image Plate set at 180 mm diameter and 150 μm pixel size. A total of 116 consecutive 1° oscillation images were collected from a single cryo-cooled crystal with a beam dosage of 3000 counts per exposure. The data were then processed

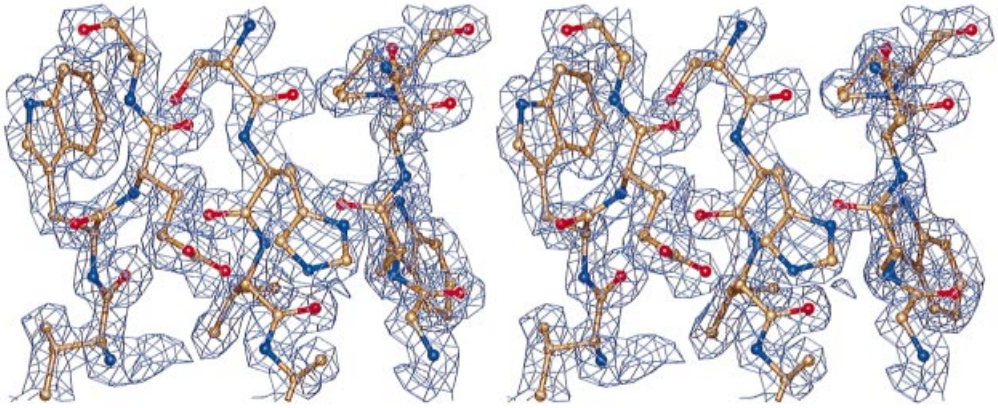


Figure 1 Electron-density around the mutated region of *T. r.* Cel7A displayed in divergent stereo centered at His^{224}

using DENZO and SCALEPACK software [22]. Data-processing statistics are given in Table 2.

X-ray crystallography: phasing and refinement

Initial phases were taken from the refined co-ordinates of *T. r.* Cel7A E217Q mutant in complex with cellohexaose and cellobiose {Protein Data Bank (PDB) accession code 7CEL; [9]} with temperature factors reset to ≈ 0.2 nm^2 (20 Å^2). The refinement program CNS [23] and modelling program O [24] were then used to refine the model. An initial cycle of simulated annealing was followed by five cycles of energy minimization and B-factor refinement. The R_{free} value (which describes how much the protein model deviates from the experimental data obtained in diffraction experiments [25,26]) was calculated at all stages from 2 % (1000) of the reflections, and rebuilding and solvent addition was judged by Sigma A weighted ($2|F_o|-|F_c|$, α_c) and ($|F_o|-|F_c|$, α_c) maps. Statistics for the final model can be found in Table 2. The electron density around the mutated residues is very clear (see Figure 1).

Kinetic measurements on soluble substrates

The liberation of 3,4-dinitrophenol (DNP) from its β-cellobioside and lactoside was measured by continuous monitoring of absorbance of solutions in 1-cm-pathlength cuvettes in a Perkin–Elmer Lambda 18 UV/VIS spectrophotometer fitted with the manufacturer's Peltier-effect thermostatically controlled cell block (25.0 °C). For k_{cat}/K_m measurements, each hydrolysis was performed in a quartz cuvette in a total volume of 700 μl. Absorbance changes were monitored at 338 nm (1 nm slit width), which was determined as the isosbestic point for 3,4-DNP and its conjugate base. Enzyme concentrations were 2.07 μM and 3.97 μM for *T. r.* Cel7A wt and mutant respectively. In order to perform the experiments with initial substrate concentration ($[S_0]$) $\ll K_m$, the K_m value for hydrolysis of 3,4-DNP-Lac by mutant Cel7A was determined (Table 3). Substrate concentrations for each experiment were 29.0 and 40.2 μM for Cel7A wt and Cel7A mutant respectively. At least 2000 data points were collected for the determination of the first-order rate constant for each experiment. Buffer systems used were sodium phosphate (pH 2–3), sodium formate (pH 3–4), sodium acetate (pH 4–5.5 for all enzymes), sodium citrate (pH 5.5–6.25) and potassium phosphate (pH 6.25–7.00). Buffer concentrations were 0.1 M in all cases. Alternative buffers have been used in overlapping pH ranges to monitor buffer and ionic-strength effects; with a couple of exceptions (below) these were not important. As a further control, doubling the concentration of formate, acetate and citrate buffers was shown to result in a less-than-6 % change in the first-order rate constants. Buffers containing an appreciable fraction of fully protonated citrate or succinate were found to be inhibitory, and were therefore not used. pH values were measured at the beginning and end of a kinetic run to check that there was no change. First-order rate constants below pH 2 for Cel7A wt and below pH 3 for Cel7A mutant were not determined because of enzyme denaturation during the experiments.

For separation of Michaelis–Menten parameters, at least ten initial-rate measurements at substrate concentrations from $0.2 \times K_m$ to $5 \times K_m$ were made. The reaction was monitored at 400 nm, where $\Delta\epsilon = 5040$ cm$^{-1} \cdot$ M^{-1} at pH 5.0 and 13 700 cm^{-1} $\cdot$ M^{-1} at pH 7.0. Time courses were fitted to an exponential (eqn 4), or initial rates to a rectangular hyperbola, using Fig P for Windows:

$$[S] = [S]_0 \cdot \exp\left[\left(-\frac{[E]_t k_{cat}}{K_m}\right) \cdot t\right] \quad (4)$$

Fitting of initial-rate data gave r^2 (square of the correlation coefficient) values > 0.93, indicating that the precision of the data collected did not limit the accuracy of the parameters in Table 3. This was more likely to be limited by weighing and dilution errors (probably < 10 %), in the case of K_m, and enzyme activity, particularly in the absence of a suitable active-site titration, in the case of k_{cat}.

Liberation of 2-chloro-4-nitrophenol (CNP) from CNP-Lac was monitored using a Bio-Rad Benchmark Easy Reader, set up for automated absorbance experiments performed in microtitre plates. For relative k_{cat} measurements, a discontinuous experiment was performed in microtitre plates with CNP-Lac as substrate. The K_m value for Cel7A wt with CNP-Lac is 460 ± 20 μM (pH 5.7, 37 °C) [17]. McIlvaine's citric acid/phosphate (0.1 M) buffers were used between pH 3.7 and 7.1. Cel7A wt (0.34 μM) and Cel7A mutant (0.52 μM) were incubated at 37 °C with 1 ml of 2 mM CNP-Lac. At regular time intervals 100 μl aliquots were mixed with 100 μl of 10 % (w/v) Na_2CO_3 to stop the reaction. The released CNP was measured at 405 nm. For measurements of inhibition by cellobiose the absorbance was monitored continuously at pH 5.7 and 33 °C. Cellobiose

concentrations up to 1.0 mM were used. Data fitting to eqn (5) below was carried out using Fig P or Mathematica [27].

Activity on crystalline cellulose

Bacterial microcrystalline cellulose (BMCC) was prepared as described previously [28]. Cellulose suspension (0.7 g/l) was shaken at 27 °C with the enzyme solution (1.4 μM) in one of the following buffers: 40 mM sodium citrate, pH 3.0, 40 mM sodium acetate, pH 5.0, or 40 mM potassium phosphate, pH 7.0. The final volume of the reaction mixture was 325 μl. The reaction was stopped at designated time points by adding half the reaction volume of stop reagent (containing 1 vol. of 1 M glycine, pH 11, and 9 vol. of 94 % ethanol) and filtering the sample through Millex GV 0.22 μm-pore-size filtration units (Millipore). The formation of reducing sugars in the supernatant was determined by the *p*-hydroxybenzoyl hydrazide method [29]. A freshly made solution (100 μl) of 0.1 M *p*-hydroxybenzoyl hydrazide (Sigma) in 0.5 M NaOH was added to 150 μl of the filtered sample and boiled for 10 min. The sample was cooled and the concentration of soluble reducing sugar was calculated from the A_{405} and a cellobiose calibration curve. HPLC analysis was performed, essentially as described earlier [30], to confirm that the main soluble product produced by both enzymes at all pH values was cellobiose. It has been shown previously [31] that *T. r.* Cel7A does not have a pronounced effect on the degree of polymerization of BMCC, and thus measuring the production of cellobiose gives a good estimate of overall hydrolytic activity of Cel7A on BMCC.

Enzyme-stability measurements

The stability of *T. r.* Cel7A wt and Cel7A mutant as a function of the pH was checked at different temperatures (25, 37 and 50 °C). The enzymes were preincubated in the buffers used for k_{cat}/K_m measurements for 1 h. Aliquots (10 μl) were taken at different time intervals and the activity was measured on 2 mM CNP-Lac (200 μl) in phosphate buffer (50 mM, pH 5.7) at 25 °C by continuously monitoring CNP release for 10 min.

Data deposition

The atomic co-ordinates and experimental structure factors have been deposited with the PDB. PDB accession codes are 1EGN and R1EGNSF respectively.

RESULTS AND DISCUSSION

Structure

For convenience, *T. r.* Cel7A numbering will be used for all structures in the following descriptions.

The electron-density maps are of very good quality (Figure 1) and indicate that the backbone structure around the modified region is essentially unchanged by the mutations (Figure 2). The root-mean-square (RMS) deviation between the wt and mutant structures is 0.11 Å, calculated for all atoms. The side chains align well with corresponding residues in the wt *H. i.* Cel7B structure (Figure 3). Ser^{223} overlaps almost perfectly with its *H. i.* Cel7B equivalent, as does Val^{225}, except for a side-chain rotamer shift of 120°. The structures show that we were correct to decrease the bulk of the residue-226 side chain, since the shortest distance between the superimposed structures is only 2.1 Å from the position of Thr^{226} in the wt enzyme to His^{224} in the mutant. However, the histidine residue might have been accommodated in the position taken on in the mutant without the replacement of Asp^{262} by glycine. The shortest distance is 3.1 Å between the residues in the superimposed structures.

The imidazole ring of His^{224} of the mutant is in plane with its counterpart in *H. i.* Cel7B, but is shifted 1.0 Å closer to the catalytic acid/base Glu^{217}. At the resolution of the present study we are unable, on the basis of the crystallographic data alone, to identify which are the nitrogen and carbon atoms in the imidazole ring. With the atomic assignment shown in Figure 4(A), the His^{224} $N^{\epsilon 2}$ atom interacts with a water molecule (at a distance of 2.8 Å) and the catalytic acid/base Glu^{217} (3.1 Å to $O^{\epsilon 1}$). The geometry for a hydrogen bond to the glutamic acid is somewhat distorted (e.g. the angle $C^{\delta 2}$–$N^{\epsilon 2}$–$O^{\epsilon 1}$ is 87°). With this conformation, His^{224} $N^{\delta 1}$ is positioned 3.2 Å from another water molecule. The alternate assignment was rejected because it would place the nitrogen atoms without hydrogen-bonding partners.

Although at one level the structures suggest that part of the catalytic machinery of *H. i.* Cel7B has been successfully incorporated into *T. r.* Cel7A, there are subtle differences in the hydrogen-bonding patterns which could influence kinetic behaviour. These differences are illustrated in Figure 4, which shows the *T. r.* Cel7A mutant (Figure 4A) and *H. i.* Cel7B (Figure 4B; PDB accession code 2A39; [13]). Also included is the structure of another Family-7 endoglucanase from *Fusarium oxysporum* in complex with a non-hydrolysable substrate analogue, a thiocellopentaoside with sulphur atoms instead of oxygen in the glycosidic bonds (Figure 4C; PDB code 1OVW; [32]). This enzyme shows high sequence similarity with *H. i.* Cel7B, also containing a histidine at position 224, and has a pH optimum around pH 7.5 [33,34]. For comparison with wt *T. r.* Cel7A, and in order to illustrate the most probable hydrogen-bonding pattern in the enzyme–substrate complex, we have chosen to use the model of *T. r.* Cel7A wt with cellononaose (Figure 4D; PDB code 8CEL; [9]), which is based on the crystal structures of catalytically impaired mutants of Cel7A in complex with cellobiose, -tetraose, -pentaose and -hexaose [9].

One of the most striking differences between wt Cel7A and the mutant is that, even when the substrate binds in the active site of the wt enzyme, the acid/base Glu^{217} can be regarded as solvated (Figure 4D). Between the cellobiosyl moiety in the product sites $+1/+2$ and the 'floor' of the active site there is a chain of four water molecules. The one closest to Glu^{217} is at a distance of 3.4 Å. The water chain is interrupted by Asp^{262}, which acts as a bridge to the surrounding solution. In the mutant the three water molecules closest to Glu^{217} are replaced by the introduced histidine residue, thereby reducing its solvation (Figure 4A).

Stability

The effect of pH on the stability of *T. r.* Cel7A wt and its mutant was checked at 25, 37 and 50 °C by incubating at pH 2.0, 3.0, 4.0, 5.0, 6.0 and 7.0 for 1 h and assaying at optimal pH and 25 °C. Both enzymes retain full activity at 25 °C at each pH. At 37 °C both enzymes were stable at pH 3–7, but at pH 2 both enzymes lost 90 % of their activity in 20 min and 100 % in 1 h. At 50 °C full activity of both enzymes is retained between pH 4 and 6, but at lower pH rapid inactivation occurs. At pH 2 there is 90 % loss of activity after 1 min and 100 % after 10 min for both enzymes, but at pH 3 the greater stability of the wt is apparent: only 30 % activity is lost from wt after 20 min, whereas the mutant completely lost activity after 10 min. 25 °C (or 27 °C) was chosen as the measurement temperature for all the activity studies of the

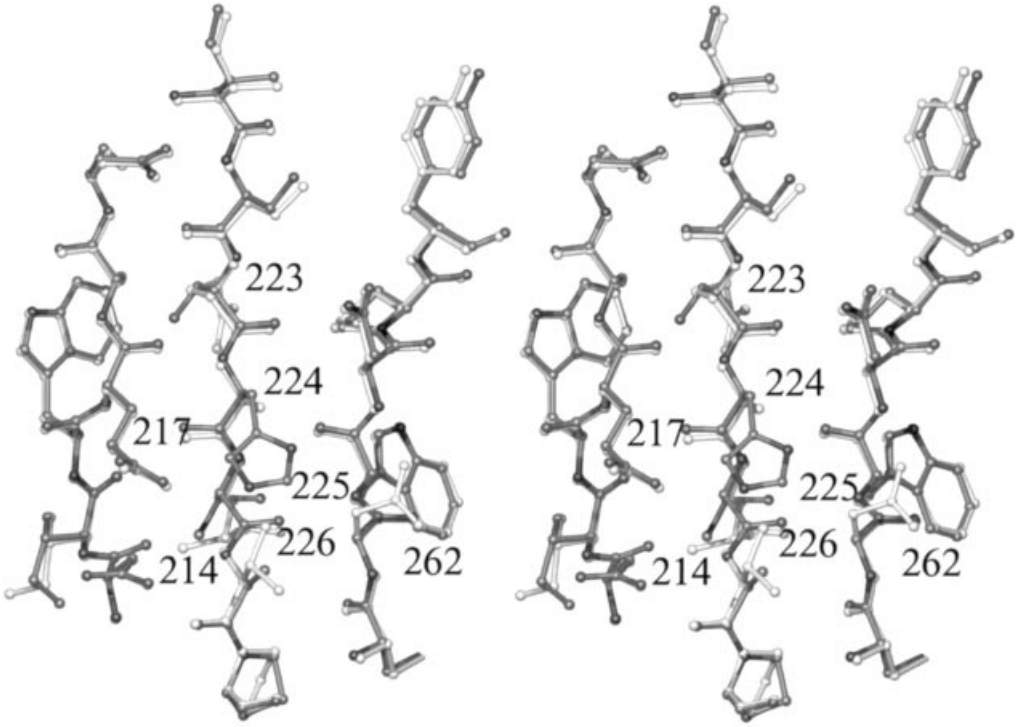


Figure 2 Comparison of *T. r.* Cel7A mutant (dark) and *T. r.* Cel7A wt (light) shown in divergent stereo

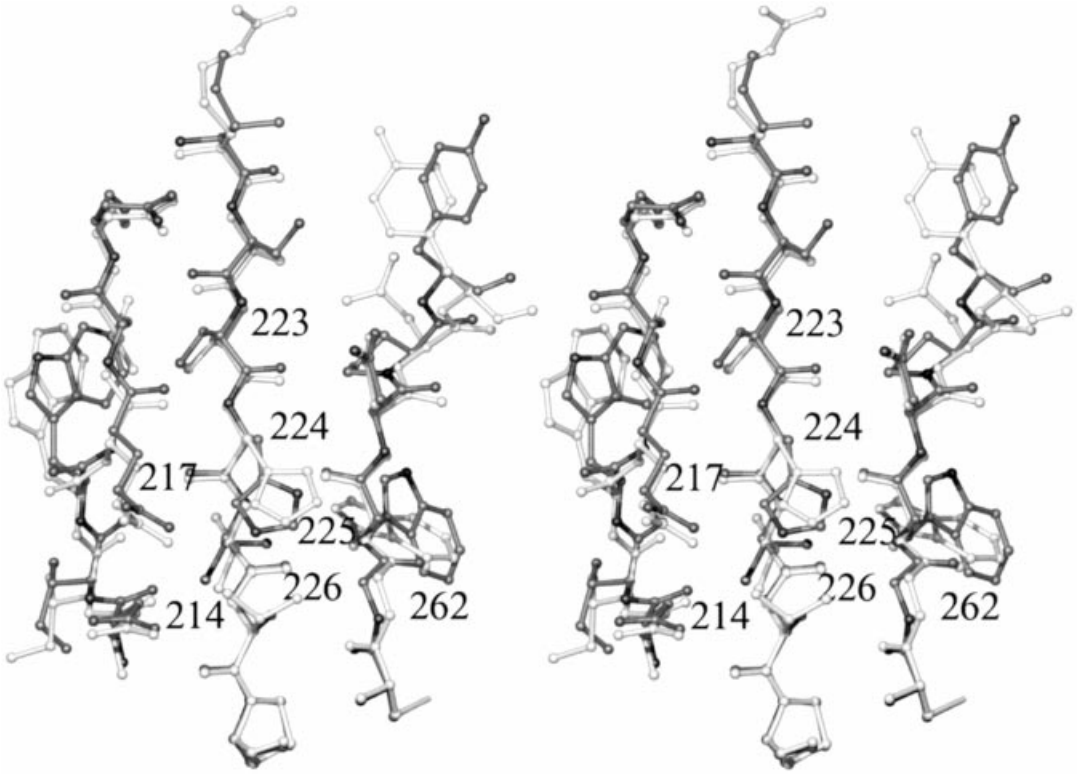


Figure 3 Comparison of *T. r.* Cel7A mutant (dark) and *H. i.* Cel7B wt (light) shown in divergent stereo

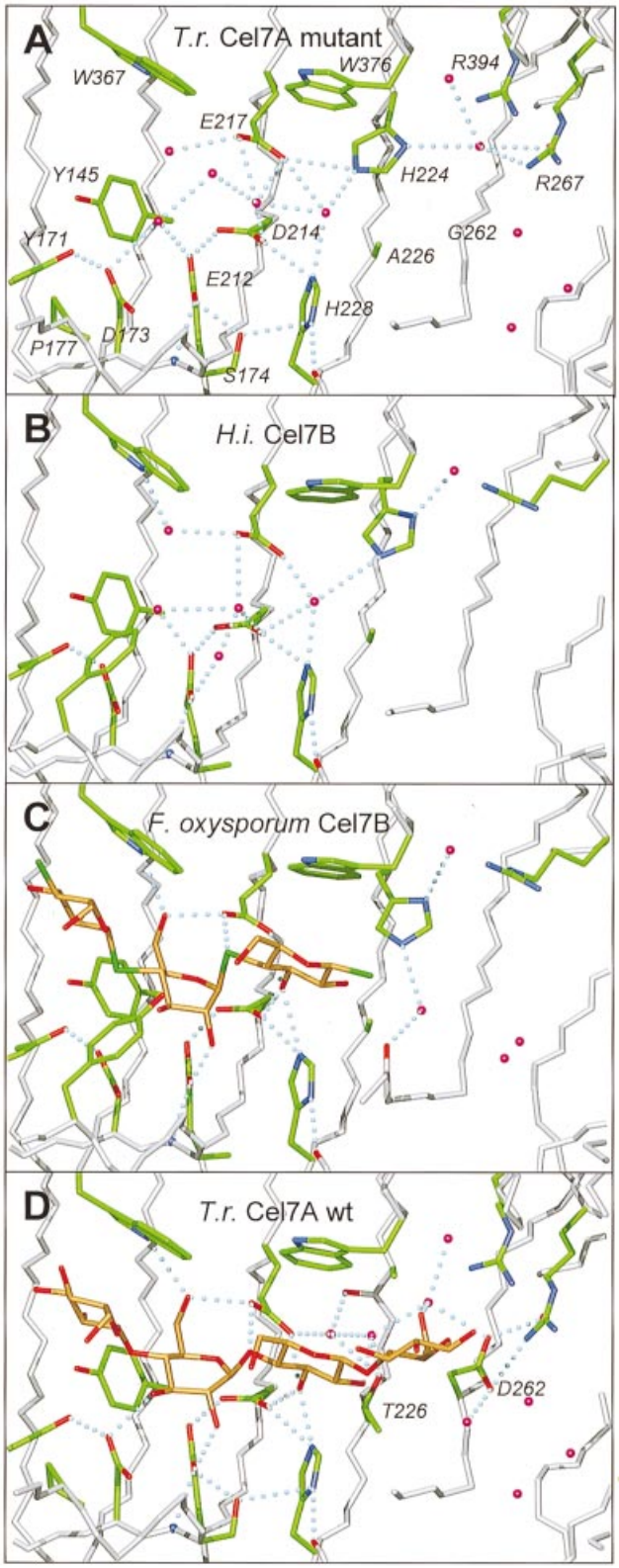


Figure 4 Comparison of interactions in the active sites of Family 7 cellulases

(**A**) *T. reesei* Cel7A mutant; (**B**) *H. insolens* Cel7B wt (PDB accession code 2A39; [13]); (**C**) *Fusarium oxysporum* endoglucanase wt in complex with thio-oligosaccharide (PDB code 1OVW; [32]); (**D**) *T. reesei* Cel7A wt with modelled cellononaose chain (PDB code 8CEL; [9]).

Cel7A wt and mutant. The lower thermal stability of the mutant may be related to a more open and flexible structure. Such changes would explain the greater sensitivity of the mutant to papain.

Kinetic effects of the mutations

Surprisingly, although pH–k_{cat} profiles for soluble substrates hydrolysed by *T. r.* Cel7A have been known for some time [35], no profile for k_{cat}/K_m, the more readily interpreted parameter, has been reported. This lacuna in the data for a well-studied enzyme possibly arose from the low K_m values for soluble substrates, which necessitated a high-precision spectrometer for data well below the K_m.

We therefore elected to use DNP-Lac as substrate. The 3,4-dinitrophenol leaving group ($pK_a = 5.4$) gives a large change in absorbance, even at slightly acid pH, and the use of the Lac, rather than the cellobioside, was dictated by the higher K_m values generally observed with aryl lactosides rather than aryl cellobiosides [36]. Moreover, as Cel7A is comparatively stable, k_{cat}/K_m was determined directly under first-order conditions with $[S]_o \ll K_m$ (and also $\ll K_i$ for the product lactose), where the eqn (4) applies.

First-order rate constants were obtained by direct fit of absorbance to an exponential for at least three half-lives; the derived total change in absorbance corresponding to complete reaction was constant within 5%, confirming that enzyme denaturation during the course of the experiment did not contribute to the derived first-order constant. The pH–rate profile for wt and the engineered mutant is given in Figure 5.

The similarity of k_{cat}/K_m values measured in different buffers at the same pH, and the absence of a buffer concentration effect, indicate that ionic-strength effects can be neglected. The data represented by the squares in Figure 5 was therefore fitted to eqn (5).

$$\log\left(\frac{k_{cat}}{K_m}\right) = \log\left[\left(\frac{k_{cat}}{K_m}\right)_{max}\right] - \log\left(1 + \frac{10^{-pH}}{K_{a_1}} + \frac{K_{a_2}}{10^{-pH}}\right) \quad (5)$$

The derived K_a values were $(6.08 \times 10^{-3}) \pm (4.3 \times 10^{-4})$ and $(1.02 \times 10^{-6}) \pm (3.9 \times 10^{-8})$ (corresponding to $pK_a = 2.22 \pm 0.03$ and $pK_a = 5.99 \pm 0.02$) for wt enzyme; K_a values were $(6.43 \times 10^{-4}) \pm (4.4 \times 10^{-5})$ and $(1.66 \times 10^{-7}) \pm (9.3 \times 10^{-9})$ (corresponding to $pK_a = 3.19 \pm 0.03$ and $pK_a = 6.78 \pm 0.02$) for the engineered mutant. The value of $\log(k_{cat}/K_m)_{max}$ is 2.89 for wt and 1.92 for the mutant. In both cases the pH-dependence of the second-order rate constant is accurately described by a classical bell shape.

If the usual assumptions of the Alberty–Bloomfield [2] treatment are made (and the low absolute values of k_{cat}/K_m make it very unlikely the substrate is 'sticky'), then these pK_a values represent ionization of the free, unliganded enzyme and are thus independent of the substrate used to determine the profile. The differences in the pH-versus-$\log(k_{cat}/K_m)$ profile between wt and engineered mutant conform to prediction in the sense that the pH optimum is more alkaline in the mutant.

Figure 6 shows the variation of relative k_{cat} for hydrolysis of CNP-Lac for wt and the mutant. The pH-dependence is broadly similar to that for k_{cat}/K_m. Since attempts by stopped-flow to observe a 'burst' of aglycone during the hydrolysis of lactosides of acidic aglycones by Cel7A fail (P. Lehtovaara, T. Selwood, M. L. Sinnott and D. W. Yates, unpublished work), k_{cat} for this substrate probably represents k_{+2}. It is clear there are no large perturbations of enzyme ionizations caused by the binding of substrate either to the wt or mutant enzyme. Fit of the data for both the wt and mutant enzyme to a bell-shaped curve gives the

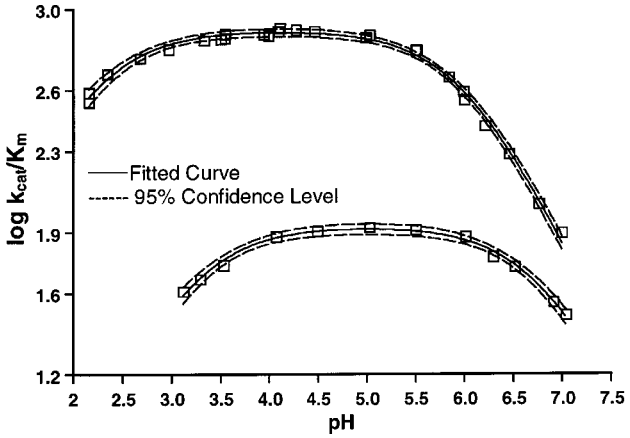


Figure 5 Variation with pH of log(k_{cat}/K_m) for hydrolysis of DNP-Lac by *T. r.* Cel7A wt (top) and its mutant (bottom), at 25 °C (95% confidence limits are drawn)

acid pK_a shifting from 2.94 ± 0.11 in the wt to 3.56 ± 0.07 in the mutant, and the basic pK_a shifting from 5.73 ± 0.04 to 6.45 ± 0.06. However, the definition of the acid ionization in the wt is not good: the fit displayed is to a single ionization of pK_a 5.82 ± 0.06.

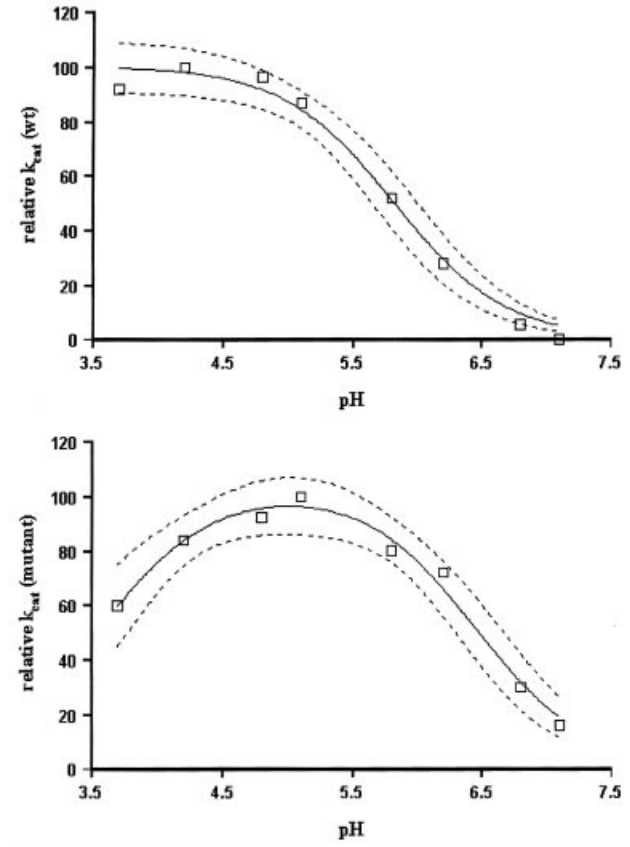


Figure 6 Variation with pH of k_{cat} (relative to the maximum) for the hydrolysis of CNP-Lac as a function of pH for *T. r.* Cel7A wt (top) and its mutant (bottom), at 37 °C (95% confidence limits are drawn)

Structural interpretation of the pH behaviour

The molecular interpretation of the shift in the pH optimum is not straightforward, because pH studies cannot distinguish between tautomers, and therefore yield macroscopic pK_a values. The active site of wt *T. r.* Cel7A, while dominated by the three acidic residues, Glu^{212}, Asp^{214} and Glu^{217}, that we have previously mutated [17], contains a complex set of interactions, some of which are indicated in Figure 4(D). The glutamyl and aspartyl side chains of amino acids 212 and 214 interact to form a close interaction (the separation $212O^{\epsilon 2}$–$214O^{\delta 1}$ varying in length from 2.5 to 2.7 Å in our different structures), which suggests that they share a proton. The carboxylate oxygen atoms involved in this close interaction do not interact with any other residues. In the unliganded enzyme they are hydrogen-bonded to water molecules, which are displaced by the glucose residue in site -1 upon formation of the enzyme–substrate (ES) complex. This leaves these carboxylate oxygen atoms in a microenvironment with no other hydrogen-bonding partners than each other. The oxygen atom Glu^{212} $O^{\epsilon 2}$ (in our numbering) is believed to perform the nucleophilic attack on the anomeric carbon.

Each of the carboxylate groups is in turn close to other acidic residues. The carboxylate oxygen atoms of Glu^{212} are 4.0–4.6 Å from the carboxylate group of Asp^{173}, which in turn forms hydrogen bonds with the main-chain nitrogen atom of residue 175 and the hydroxy group of Tyr^{171}. The carboxylate oxygen atoms of Asp^{214} are close to the carboxylate oxygen atoms of the acid/base Glu^{217} (separations of 3.5–4 Å). Furthermore, Asp^{214} $O^{\delta 2}$ interacts with the imidazole ring of His^{228} (a separation $214O^{\delta 2}$–$228N^{\epsilon 2}$ of 3.0 Å). The carboxylate oxygen atom $O^{\epsilon 1}$ of Glu^{212} makes two more hydrogen-bond contacts (hydroxy group of Ser^{174}, 2.7 Å, and the main-chain nitrogen atom of 174, 3.0 Å). In the wt, ligand-free structure, Glu^{217} is solvated and makes no hydrogen-bond interactions with the rest of the enzyme. In the mutant, an additional imidazole ring enters into the picture, close to Glu^{217}, as described above. A specific set of hydrogen-bond interactions sometimes allows one to assign which atoms are acceptors and donors. Unfortunately, in the present study we are only able to state that Glu^{212} and Asp^{214} carboxylate groups share a hydrogen bond because of their close contact. Estimating the ionization states of the other residues entails much guesswork. We can be fairly sure that the close interaction between Glu^{212} and Asp^{214} residues will result in a low first pK_a, and a much higher second value. If we imagine that the neighbouring residues influence the position of the shared proton, the interactions of either amino acid 212 or 214 with their neighbours can stabilize a charge on either residue. The accessibility of Glu^{217} to the solvent, and a fairly close contact to Asp^{214} suggests a normal or somewhat elevated pK_a. The crystallographic and mutational studies on Cel7A [9,17] and similar structural studies on the related endoglucanases [13,32], suggest that residue 212 is the nucleophile, while amino acid 217 is the acid/base. The active state of the enzyme is, therefore, expected to be: amino acid 212 charged, amino acids 214 and 217 protonated. The observed pK_a of 2.22 at the acid limb of the reaction, Figure 5, may correspond to the first ionization state of the interacting 212/214 residues, while the pK_a of 5.99 at the alkaline limb may correspond to Glu^{217}.

The behaviour of the mutant is more difficult to describe, even qualitatively. If we limit ourselves to the four acidic groups and two histidine residues, we have potentially 64 different tautomers to consider. The closest approach of the imidazole ring of 224 is clearly to amino acid 217 (≈ 3.1 Å), but perturbations may arise to Asp^{214} (and hence amino acid 212) and His^{228}, with separations of ≈ 5 Å. For example, if we assume His^{224} is charged, Glu^{217}

may prefer to be ionized. Residue 214 would then be under the opposing perturbations of residues 217 and 224. Similarly, if we suppose His^{224} is uncharged, its greatest effect on Glu^{217} may be the reduction in the solvation of the carboxylate group. Such a change may actually cause an increase in the pK_a of Glu^{217}. Experimentally, we observe shifts in the macroscopic pK_a values to 3.19 and 6.78, and a reduction in overall activity. The shift in pK_a in the acid region is probably the result of the charged histidine residue interacting with the whole set of interacting acidic groups, while the pK_a shift of the alkaline limb may be the influence of the non-charged histidine residue on the acid/base Glu^{217}. The proximity of the histidine residue may affect the equilibrium of the proton shared by amino acids 212/214, but the close interaction between these groups is clearly maintained in the crystal structure. Introducing a new positively charged residue close to the expected build-up of positive charge on O^5 and C^1 of the sugar residue in the -1 subsite in the transition state (nearest approach of 6.6 Å to C^1 in our current model for cellulose binding [9]) is also a potential cause for reduced activity.

His^{224} and Glu^{217} are likely to affect each other's microscopic pK_a values via the mutual stabilization of their ionized states. The dominant tautomer may be the inactive form of the enzyme, with both His^{224} and Glu^{217} charged, whereas the active form may constitute a minor tautomer of the same overall ionization state. Dependence of activity on pH cannot in principle distinguish between reaction through different rapidly equilibrating tautomers of the same protonation state. Macroscopic, not microscopic, pK_a values are derived from pH–rate profiles. Therefore, a bell-shaped pH profile is mathematically consistent with reaction through the minor tautomer of the dominant protonation state, which has the more acidic group protonated and the less acidic one deprotonated. Such 'reverse-protonation' mechanisms have the in-built inefficiency of using a minor tautomer, but one has been proposed for thermolysin [37]. Action via a minor tautomer has also been proposed for the xylanases of glycosidase Family 11 with acidic pH optima, which have an aspartic acid residue hydrogen-bonded to the acid/base catalyst [38].

Comparison of *T. r.* Cel7A mutant and *H. i.* Cel7B structures

Although the structures of the active sites in *H. i.* Cel7B and *T. r.* Cel7A mutant are very similar, there are some potentially important differences (Figure 4B). The histidine residue equivalent to position 224 is further away from the acid/base catalyst (closest approach is 3.7 Å), and from the residue equivalent to 214 (closest approach 6.0 Å) compared with our Cel7A mutant structure. In the *H. i.* endoglucanase, there are no close contacts to the imidazole ring. The specific interactions made by the nucleophile also change, due to the replacement of residue Ser^{174} by an alanine, resulting in the loss of a hydrogen bond. The close interaction between amino acids 212 and 214, however, is kept. The hydrogen bonds made by Asp^{173} are unchanged, but the sequence change Pro^{177} to Phe introduces the bulky aromatic ring (closest approach 4.3 Å). On the other side of the catalytic residues, in the product-binding sites, *T. r.* Cel7A contains some additional ionizable residues which are missing or replaced by non-ionizable residues in the endoglucanases (Figure 4). Out of three arginine residues in *T. r.* Cel7A at positions 251, 267 and 394, only the one at 267 is present in *H. i.* Cel7B. It is closer to His^{224} in *H. i.* Cel7B with a distance of 4.8 Å ($His\ N^{\delta 1}$ to $Arg\ N^{\epsilon}$) compared with 6.2 Å from Arg^{267} and 7.0 Å from Arg^{394} in the *T. r.* Cel7A mutant. However, the arginine side-chain is pointing away from the histidine residue, whereas in *T. r.* Cel7A both Arg^{394} and Arg^{267} are pointing into the active-site cleft. Furthermore, Glu^{334} (7 Å from His^{224}) is replaced by Thr in *H. i.* Cel7B. As a consequence, the environment of the histidine is less polar and less basic in the endoglucanases, which may influence the position of the histidine and/or contribute to the elevation of the pH optimum. The most important difference may concern the separation of the imidazole ring from the acid/base catalyst. When charged, the more distant imidazole group probably has a less profound affect on the pK_a of the acid catalyst, but when uncharged it may still assist in elevating the pK_a, due to solvation effects.

There is no obvious structural reason why the position of the histidine residue differs. In the plane of the imidazole ring there is space for sideways movement, so that, via small rotations of the side-chain torsion angles, the respective rings can be superimposed. The imidazole rings of both endoglucanases can be shifted into close contact with the acid/base Glu^{217} without creating any unfavourable contacts. The position of His^{224} in the *T. r.* Cel7A mutant may not be governed primarily by the interaction with Glu^{217}, but may be influenced by other factors. For example, electrostatic repulsion between a charged histidine residue and the three arginine residues mentioned above could be one factor, or slightly different local backbone conformations could be another. If the side-chain conformer of *T. r.* His^{224} is changed to superimpose it on the *F. oxysporum* endoglucanase, a close contact develops with Gly^{226} ($C^{\epsilon 1}$–C^{α} contact of 3.0 Å). This can be relieved by rather small backbone adjustments, which suggests that the histidine is not held in place by strong forces, but perhaps may shift position, for example upon substrate binding. Such movement would be sterically hindered if we had kept the aspartic acid side chain at position 262.

Other structural effects of the mutations

The conformation of the backbone at His^{224} is likely to be influenced by the adjacent residues, which were therefore mutated to be the same as in *H. i.* Cel7B, Glu^{223} to serine and Leu^{225} to valine, in order to facilitate possible backbone adjustments necessary to accommodate the introduced histidine. There is actually a small shift of the C^{α} atom at position 225 in our model about 0.4 Å upwards into the cleft, but we are uncertain if this is significant. Otherwise there are no changes at all compared with other *T. r.* Cel7A structures apart from atoms of mutated side chains. We cannot be sure that these mutations were necessary.

The residues at 223 and 225 point into a small cavity in the β-sheet interface underneath the active site. The cavity holds two water molecules and is similar in size in *T. r.* Cel7A and in the endoglucanases. However, the residues that line the cavity, emerging from the inner and outer β-strands, differ between the enzymes. Therefore the mutated residues will not function equivalently even if they are now identical. The reduction in size of the side chains makes the cavity bigger in the *T. r.* Cel7A mutant and affects the β-interface interactions. This may have a destabilizing effect on the protein fold and could explain why the mutant is degraded by papain in hours under conditions where the wt enzyme is resistant for several days. It may also contribute to the reduction in activity for the mutant enzyme.

Activity towards 3,4-DNP-cellobiose

An unexpected feature of the mutant emerged when activity (i.e. k_{cat}) was measured against cellobiosyl derivatives: the mutant appeared to have the higher activity. The behaviour of 3,4-DNP cellobioside was examined in more detail (Table 3): the second-order rate constant, k_{cat}/K_m, showed the same pattern for the

Table 3 Kinetic parameters for hydrolysis of chromophoric substrates by *T. r.* Cel7A wt and its quintuple mutant at 25.0 °C

(a)

	$10^3 \times k_{cat}$ (s^{-1})	
Substrate ...	3,4-DNP-cellobioside	3,4-DNP-Lac
wt		
pH 5.0	4.9	650
pH 7.0	0.42	–
Mutant		
pH 5.0	6.5	52
pH 7.0	5.4	–

(b)

	K_m (μM)	
Substrate ...	3,4-DNP-cellobioside	3,4-DNP-Lac
wt		
pH 5.0	14	380
pH 7.0	11	–
Mutant		
pH 5.0	48	543
pH 7.0	161	–

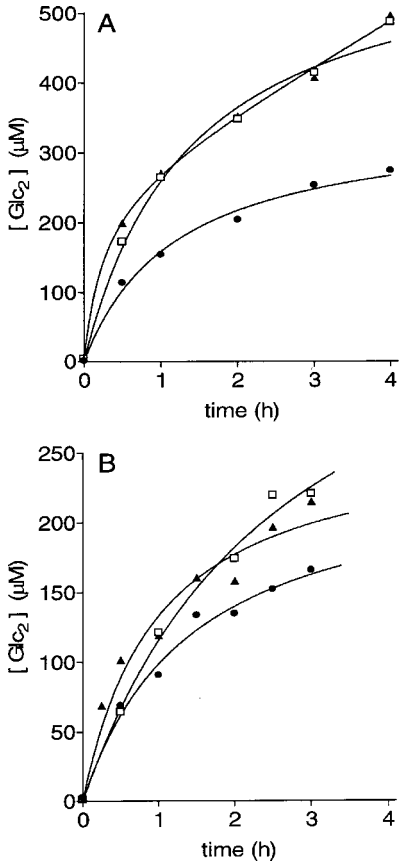


Figure 7 Hydrolysis of BMCC at 27 °C by *T. r.* Cel7A wt (A) and its quintuple mutant (B) at pH 3.0 (▲), pH 5.0 (□) and pH 7.0 (●)

The soluble products were measured at five to seven different time points as duplicates. The error at each time point is about 10–20% (not shown in the Figure).

cellobioside as for the lactoside, but the mutant showed a slight increase in k_{cat} and a more-than-compensating increase in K_m. A long-standing puzzle about *T. r.* Cel7A is the higher k_{cat} values for lactosides compared with cellobiosides. With the X-ray data now available for *T. r.* Cel7A binding small ligands [9,17], it is clear that the cellobiosyl moiety directs binding in the first instance to the +1, +2 sites. The structural expectation, then, is that most of the ES complex of *T. r.* Cel7A with an aryl cellobioside will be in the form of a non-productive complex with the glucosyl residues in the +1, +2 sites. The effect of non-productive binding of this type on steady-state kinetics is to divide the expressions both for k_{cat} and K_m (eqns 1 and 2) through by terms of the form $(1+K_s/K_{np})$, where K_{np} is the dissociation constant of the complex in the non-productive binding mode, and K_s refers only to the productive binding mode. Consequently, the effect of non-productive binding is to reduce k_{cat} and K_m equally and leave k_{cat}/K_m unchanged. The relative activities of *T. r.* Cel7A wt on cellobiosides and lactosides are therefore consistent with cellobiosides adopting a non-productive (+1, +2) binding mode which is disfavoured in the case of lactosides by the axial OH group of the galactose residue. In addition, the results on the Cel7A mutant show a weakening of the inhibition by cellobiose. The inhibition is strictly competitive with the wt enzyme, for which $K_I = 20$ μM [39]. However, with the mutant the inhibition is much weaker and mixed, with $K_I = 755$ μM, α (ratio of anticompetitive to competitive binding constants) = 2.9. Thus the five point mutations in *T. r.* Cel7A, besides changing to the pH optimum of the enzyme, also selectively weaken the binding of glucose units at the +1, +2 sites and lead to apparently higher activity on cellobiosides.

Structurally the mutations do not alter the shape of the product site or the residues that interact directly with the cellobiose moiety. Rather the weaker binding is due to the loss of two water-mediated hydrogen bonds present in wt *T. r.* Cel7A, one from the Thr^{226} side chain via a water molecule to OH3 of the glucosyl residue in site +1, and the other from the Asp^{262} side chain via water to OH6 of the glucosyl in site +2 (Figure 4D). Both these residues were mutated, residue 226 to Ala and residue 262 to Gly, and, furthermore, the two water molecules are displaced by the imidazole side chain of His^{224} in the mutated enzyme (Figure 4A).

Activity towards crystalline cellulose

The activity of the mutant towards solid BMCC (Figure 7) mirrors the activity against 3,4-DNP-Lac, with the activity about 20–30% of wt (calculated as the time that it takes for the wt and the mutant enzyme to produce the same amount of cellobiose) at pH 5, increasing to 40–50% at pH 7. However, the hydrolysis rate of the mutant enzyme at pH 3 on BMCC seems to be similar to that at pH 5, which is in contrast with lactoside hydrolysis. Given that the rate-determining step in the hydrolysis of solid substrates is likely to be physical, the existence of differences in the effect of the mutation on solid substrates and on lactosides is unremarkable.

Concluding remarks

The shift to a more alkaline pH optimum of the *T. r.* Cel7A mutant verifies our hypothesis that the presence of the histidine

adjacent to the acid/base catalyst contributes to the higher pH optima in the *H. i.* and *F. oxysporum* endoglucanases. Although the observed positions of the imidazole side chain differ, we believe that it affects the ionization properties of the catalytic machinery in the same way in the Cel7A mutant and in the endoglucanases, possibly by decreasing the solvation of the catalytic acid/base. Interestingly, the mutation of a histidine (His^{86}) positioned at a distance of 4.4 Å from the acid/base catalyst (Glu^{128}) in the structurally unrelated Family-10 xylanase XynA from *Streptomyces lividans*, had the opposite effect on the pH profile [40]. All of the six substitutions made (to Phe, Trp, Ala, Glu, Gln or Lys) caused a decrease in the pK_a at the alkaline limb by 0.5 to 1.9 units from the value of 9.4 in the wt enzyme.

Our mutations shifted the *T. r.* Cel7A enzyme towards the alkaline behaviour of the endoglucanases, but not all the way. We do not know if that is because the modified catalytic machinery requires further adaptations in terms of creating the proper local backbone conformation and dynamics around the histidine residue and/or establishing the correct electrostatic microenvironment for it. The pH profile may be as strongly influenced by differences in the distribution of charged residues around the active site as well as the overall charge of the enzyme. Indeed the pI values for the enzymes reflect their pH optima. The theoretical pI values estimated from the amino acid compositions are for *T. r.* Cel7A wt 4.3, *T. r.* Cel7A mutant 4.4, *H. i.* Cel7B 5.8 and *F. oxysporum* endoglucanase 7.9 (calculated using DNA-Star). The same general trend, i.e. that enzymes with acidic pH optima have low pI values, whereas alkaline enzymes have high pI values, have been reported to occur among the retaining xylanases in glycoside hydrolase Family 11 [38].

We thank Sanna Auer and Tiina Kinnari (VTT Biotechnology) for their help in the *Trichoderma* transformation, cultivations and/or protein characterization work. Tiina Liljankoski, Kati Ruotsalainen, Riitta Suihkonen (VTT) and Sirkka Kanervo (Röhm Enzyme Finland OY) are thanked for skilful technical assistance; Matti Siika-aho (VTT) for providing Cel7A wt enzyme preparations, and Tarmo Pellikka (VTT) for help with HPLC analysis. This work was financed by grant BIO4-CT96-0580 from the Biotechnology Programme of the European Union (all laboratories), the Paper Federation of Great Britain (UMIST), the Swedish Foundation for Strategic Research via the Centre for Forest Biotechnology and Chemistry, Bo Rydins Foundation for Scientific Research and the Swedish Council for Forestry and Agricultural Research (Uppsala).

REFERENCES

1 Davies, G., Sinnott, M. L. and Withers, S. G. (1998) Glycosyl transfer. In Comprehensive Biological Catalysis (Sinnott, M. L., ed.), vol. 1, pp. 119–208, Academic Press, London
2 Alberty, R. A. and Bloomfield, V. (1963) Multiple intermediates in steady state kinetics. V. Effect of pH on the rate of a simple enzymatic reaction. J. Biol. Chem. **238**, 2804–2810
3 Béguin, P. and Lemaire, M. (1996) The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. CRC Crit. Rev. Biochem. Mol. Biol. **31**, 201–236
4 Henrissat, B., Teeri, T. T. and Warren, R. A. (1998) A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. FEBS Lett. **425**, 352–354
5 Claeyssens, M., Nerinckx, W. and Piens, K. (eds.) (1998) Carbohydrases from *Trichoderma reesei* and Other Microorganisms, Royal Society of Chemistry, Cambridge
6 Henrissat, B. and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. **316**, 695–696
7 Henrissat, B. and Davies, G. J. (1997) Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. **7**, 637–644
8 Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T. and Jones, T. A. (1994) The three-dimensional structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science **265**, 524–528
9 Divne, C., Ståhlberg, J., Teeri, T. T. and Jones, T. A. (1998) High resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. J. Mol. Biol. **275**, 309–325
10 Barr, B. K., Hsieh, Y.-L., Ganem, B. and Wilson, D. B. (1996) Identification of two functionally distinct classes of exocellulases. Biochemistry **35**, 586–592
11 Boisset, C., Fraschini, C., Schülein, M., Henrissat, B. and Chanzy, H. (2000) Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. Appl. Environ. Microbiol. **66**, 1444–1452
12 Kleywegt, G. J., Zou, J.-Y., Divne, C., Davies, G. J., Sinning, I., Ståhlberg, J., Reinikainen, T., Srisodsuk, M., Teeri, T. T. and Jones, T. A. (1997) The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. J. Mol. Biol. **272**, 383–397
13 MacKenzie, L. F., Sulzenbacher, G., Divne, C., Jones, T. A., Wöldike, H. F., Schülein, M., Withers, S. G. and Davies, G. J. (1998) Crystal structure of the family 7 endoglucanase I (Cel7B) from *Humicola insolens* at 2.2 Å resolution and identification of the catalytic nucleophile by trapping of the covalent glycosyl-enzyme intermediate. Biochem. J. **335**, 409–416
14 Schülein, M. (1997) Enzymatic properties of cellulases from *Humicola insolens*. J. Biotech. **57**, 71–81
15 Margolles-Clark, E., Hayes, C. K., Harman, G. E. and Penttilä, M. (1996) Improved production of *Trichoderma harzianum* endochitinase by expression in *Trichoderma reesei* Appl. Environ. Microbiol. **62**, 2145–2151
16 Mach, R. L., Schindler, M. and Kubicek, C. P. (1994) Transformation of *Trichoderma reesei* based on hygromycin B resistance using homologous expressional signals. Curr. Genet. **25**, 567–570
17 Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeyssens, M., Teeri, T. T. and Jones, T. A. (1996) Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. J. Mol. Biol. **264**, 337–349
18 Tomme, P., van Tilbeurgh, H., Pettersson, G., van Damme, J., Vanderkerckhove, J., Knowles, J., Teeri, T. and Claeyssens, M. (1988) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur. J. Biochem. **170**, 575–581
19 Koivula, A., Lappalainen, A., Virtanen, S., Mäntylä, A. L., Suominen, P. and Teeri, T. T. (1996) Immunoaffinity chromatographic purification of cellobiohydrolase II mutants from recombinant trichoderma reesei strains devoid of major endoglucanase genes. Protein Expression Purif. **8**, 391–400
20 van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, T. and Pettersson, G. (1986) Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. Separation of functional domains. FEBS Lett. **204**, 223–227
21 McPherson, A. (1982) Preparation and Analysis of Protein Crystals. Wiley, New York
22 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276**, 307–326
23 Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S. et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. Sect. D **54**, 905–921
24 Jones, T. A., Zou, J.-Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. Sect. A **47**, 110–119
25 Brünger, A. T. (1992) Free *R* value: a novel statistical quantity for assess the accuracy of crystal structures. Nature (London) **355**, 472–475
26 Kleywegt, G. J. and Brünger, A. T. (1996) Checking your imagination: applications of the free *R* value. Structure **4**, 897–904
27 Wolfram, S. (1996) The Mathematica Book, 3rd edn. Wolfram Media/University Press, Cambridge
28 Gilkes, N. R., Jervis, E., Henrissat, B., Tekant, B., Miller, Jr., R. C., Warren, R. A. J. and Kilburn, D. G. (1992) The adsorption of a bacterial cellulase and its isolated domains to crystalline cellulose. J. Biol. Chem. **267**, 6743–6749
29 Lever, M. (1972) A new reaction for colorimetric determination of carbohydrates. Anal. Biochem. **47**, 273–279
30 Teleman, A., Koivula, A., Reinikainen, T., Valkeajärvi, A., Teeri, T. T., Drakenberg, T. and Teleman, O. (1995) Progress curve analysis shows that glucose inhibits the cellotriose hydrolysis catalysed by cellobiohydrolase II from *Trichoderma reesei*. Eur. J. Biochem. **231**, 250–258
31 Srisodsuk, M., Kleman-Leyer, K., Keränen, S., Kirk, T. K. and Teeri, T. T. (1998) Modes of action on cotton and bacterial cellulose of a homologous endoglucanase–exoglucanase pair from *Trichoderma reesei*. Eur. J. Biochem. **251**, 885–892
32 Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M. and Davies, G. J. (1996) Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group. Biochemistry **35**, 15280–15287

33 Schülein, M., Kauppinen, M. S., Lange, L., Lassen, S. F., Andersen, L. N., Klysner, S. and Nielsen, J. B. (1998) Characterisation of fungal cellulases for fibre modification. In Enzyme Applications in Fibre Processing (Eriksson, K.-E. L. and Cavaco-Paulo, A., eds.), ACS Symposium Series vol. 687, pp. 66–74, American Chemical Society, Columbus, OH
34 Mackenzie, L. F., Davies, G. J., Schülein, M. and Withers, S. G. (1997) Identification of the catalytic nucleophile of endoglucanase I from *Fusarium oxysporum* by mass spectrometry. Biochemistry **36**, 5893–5901
35 van Tilbeurgh, H. (1986) Studie van het cellulasecomplex uit *Trichoderma reesei*. Ph.D. Thesis, University of Ghent
36 Claeyssens, M., van Tilbeurgh, H., Tomme, P., Wood, T. M. and McRae, S. I. (1989) Fungal cellulase systems. Comparison of the specificities of the cellobiohydrolases isolated from *Penicillium pinophilum* and *Trichoderma reesei*. Biochem. J. **261**, 819–825

Received 13 July 2000/26 January 2001; accepted 2 March 2001

37 Mock, W. L. and Stanford, D. J. (1996) Arazoformyl dipeptide substrates for thermolysin. Confirmation of a reverse protonation catalytic mechanism. Biochemistry **35**, 7369–7377
38 Joshi, M. D., Sidhu, G., Pot, I., Brayer, G. D., Withers, S. G. and McIntosh, L. P. (2000) Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. J. Mol. Biol. **299**, 255–279
39 Vonhoff, S., Piens, K., Pipelier, M., Braet, C., Claeyssens, M. and Vasella, A. (1999) Inhibition of cellobiohydrolases from *Trichoderma reesei*. Synthesis and evaluation of some glucose-, cellobiose-, and cellotriose-derived hydroximolactams and imidazoles. Helv. Chim. Acta **82**, 963–980
40 Roberge, M., Shareck, F., Morosoli, R., Kluepfel, D. and Dupont, C. (1998) Site-directed mutagenesis study of a conserved residue in family 10 glycanases: histidine 86 of xylanase A from *Streptomyces lividans*. Protein Eng. **11**, 399–404