Cellulose, a linear, essentially insoluble \(\beta\)-1,4-glucosidically linked polymer with a size of many thousand glucose units, is one of the main components of plant materials. Many micro-organisms can degrade this polymer; they produce many different cellulases which differ in both their mode of action and their overall folds. Fungal organisms which degrade cellulose produce many different cellulases which can be classified into families according to their sequences [13]. The saprophytic fungus *Humicola insolens* produces seven different cellulases which belong to five different families (family 5, 6, 7, 12, and 45); it produces one endoglucanase (EG) from each family and two cellobiohydrolases (CBH) from two different families (6 and 7) [2,3]. One of the best characterized cellulolytic fungi, *Trichoderma reesei*, produces at least six cellulases [4].

The catalytic activity of cellulases can be determined by many different methods. However, the real substrate, cellulose, is heterogeneous and contains both amorphous and crystalline areas; the cellulases will have different reaction rates depending on which part of the structure they degrade. The cellulose binding domain (CBD) present in many cellulases can also result in non-productive binding. For accurate kinetic determinations, the ideal substrate should be soluble; for enzymatic characterization of fungal cellulases, soluble cellohexaotol and carboxymethyl cellulose (CMC) have been used. For obtaining information on the apparent kinetics of an insoluble substrate, the insoluble phosphoric acid swollen cellulose (PASC) was chosen as the most amorphous and homogeneous cellulose.

The activity of the cellulase on the different substrates was assessed by measurement of the formation of reducing end groups after 20 min incubation using the PHBAH method modified from Lever [5]. Alternative *in situ* steady state kinetics were obtained by using a coupled assay using an oxidoreductase enzyme and an electron acceptor chromophore [2,3].

Cellulases can be cloned by expression cloning in yeast as described [6]. An alternative cloning strategy is to use molecular cloning with PCR techniques and use the best conserved amino acid residues for construction of primers [7]. Using both methods family 45-homologous cellulases have been cloned recently from *Acremonium* sp., *Aspergillus aculeatus*, *Crinipellis scapula*, *Fusarium oxysporium*, *Macrophomina phaseola* and *Myceliophthora thermophila*.

The family 45 fungal cellulases are characterized by a catalytic domain of 202 and 214 amino acid residues which fold as *Humicola insolens* EG V (Cel45) [8]. The native structure of *Humicola insolens* Cel45 was solved in 1993 [8], and later, three other structures were solved, two of which contained substrate bound in the active site [9]. The family 45 endoglucanases consist of an active site which is a long shallow groove about 40 Å long, giving space for seven glucose units (subsite \(-4\) to \(+3\)). During catalysis the groove becomes covered by a flexible loop. The kinetic data obtained with six family 45 cellulases (described above) indicate that they have the highest activity on PASC [7]. Using a shorter substrate such as cellohexaotol, the catalytic rate is reduced (because the seven subsites are not covered by this substrate) [7].

The catalytic core of family 45 cellulases is linked to a CBD at the C-terminus via an O-glycosylated linker region. The structure and flexibility of the linker region is not known, but all fungal CBDs have high sequence similarity indicating the same three-dimensional structure as the CBD from *Trichoderma reesei* CBHI.
whose three-dimensional structure was solved in 1989 by Kraulis et al. [10].

Using a long soluble substrate like CMC, which has a degree of substitution of 0.7 and a degree of polymerization (DP) of approximately 200 (like PASC), the catalytic rate is also slightly reduced. This could be due to steric hindrance due to the substitutions [7].

All of the cloned cellulases were produced by cloning the cDNA into a plasmid with cDNA inserted between the fungal amylase promoter and AMG terminator from *Aspergillus niger*, followed by transformation into *Aspergillus oryzae* [11]. The cellulases with CBD were purified using Avicel affinity chromatography and the EGs without CBDs were purified using cation exchange chromatography as described for *Humicola insolens* cellulases [3].

All cellulases studied have been sequenced and characterized. They are all produced in gram amounts using an *Aspergillus oryzae* expression system. We have developed different steady-state kinetic systems for obtaining information about the catalytic properties of those different cellulases.

The family 45 cellulase pH activity profiles were obtained using CMC substrate and the results are shown in Figure 1. *Crinipellis* Cel45 has the best activity at low pH (5–6). The *Thielavia* and *Myceliophthora* Cel45 have maximal activity at neutral pH (6–7.5) and *Acremonium*, *Machrophomina* and *Humicola* have the best alkaline activity with optimum at pH 7.5–8.3. The pH activity profile was obtained using 5–10 times the apparent $K_M$ for CMC (7.5 g/l). The apparent CMC kinetics measured at pH 5.5 and 7.5 gave a $K_M$ between 0.5 and 1.5 g/l for all six cellulases [7]. The apparent $k_{cat}$ on CMC was highest for *Humicola* and *Acremonium* Cel45 at pH 7.5 and highest for *Crinipellis* Cel45 at pH 5.5 and about equal at both pH values for the others [7] in agreement with the profile in Figure 1.

The family 45 cellulase sequences show between 65 and 80% homology with the known *Humicola* sequence [7]. Of the 21 amino acids which interact directly or indirectly via water with cellohexaose [9] 20 are structurally conserved in all five of the new family 45 cellulases. This indicates that pH dependence must be due to amino acids not directly involved in the binding of substrate during catalysis. Other regions of the enzyme molecule are therefore responsible for establishing the pH optima. Work is in progress using site-directed mutagenesis to identify those regions of this family of cellulases.

Table 1 summarizes one attempt to understand the mechanism of one of the subsites in more detail. Tyrosine-147 in *Humicola insolens* Cel45 is conserved in all of the cloned family 45 cellulases and it defines subsite $+1$. This residue was substituted with 14 other amino acids and the variants have been purified and characterized. The catalytic activity was determined using reduced cellohexaose at pH 7.5 and 9.5 and the calculated $k_{cat}$ was compared with the wild-type enzyme. The catalytic activity on PASC was also assessed at pH 8.5 and 10. The data indicate that Tyr-147 can be replaced with another aromatic amino acid or a methionine residue with-

\[\text{Figure 1}\]

**Family 45 endoglucanase pH activity profiles**

CMC substrate was used at 7.5 g/l and incubation was at 40°C for 20 min at the different pH values shown. The relative activity is calculated as a percentage of the highest activity obtained.
out much adverse effect. The other substitutions resulted in a pronounced reduction of activity, particularly the alkaline activity. For the charged and the hydrophilic amino acids, the reduction in alkaline activity was most pronounced using the shorter substrate cellobetaltiol, which does not cover the whole active site groove, indicating that the contribution of more subsites will reduce the effect of one subsite.

Another family of fungal cellulases which is of commercial interest is family 7. In this family of cellulases, only fungal genes have so far been identified and in *Humicola insolens* they code for an EG formerly known as EGI (now called Cel7B) and for the cellobiohydrolase I (now called Cel7A). In this family the structure of the *Trichoderma reesei* CBHI catalytic domain as well as the *Fusarium* and *Trichoderma reesei* EGI catalytic domain have been solved [12–16]

The pH activity of different EGs from this family was determined using CMC as substrate (Figure 2). *Trichoderma reesei* EGI has a linker and a CBD and the enzyme was purified as described in [17]. The three other Cel7Bs did not have a CBD, they were all cloned and expressed in *Aspergillus oryzae* and purified as described in [3]. The endoglucanase from *Trichoderma* and *Myceliophthora* show pH optima around 4.5–5.5 whereas *Fusarium* and *Humicola* Cel7B both have higher activity at pH 8. The *Kₘ* of family 7 endoglucanases against CMC is between 1.5 and 13 g/l meaning that in this case the *Kₘ* is close to the concentration used for the activity profile. However, the CMC solutions are so viscous at higher concentrations that a diffusion problem arises.

The sequence alignment of the different family 7 EGs shows high homology particularly at those amino acids directly involved in catalysis as seen from the three-dimensional structures with sugars or sugar analogues in the active site [13–15]. Site-directed amino acid substitution has also been done for the *Humicola insolens* family 7 endoglucanase [18].

In conclusion, we have obtained a lot of

### Table 1

<table>
<thead>
<tr>
<th>Amino acids</th>
<th>pH 7.5</th>
<th>pH 9.5</th>
<th>pH 8.5</th>
<th>pH 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Red DP 6</td>
<td>Red DP 6</td>
<td>PASC</td>
<td>PASC</td>
</tr>
<tr>
<td>Y147W</td>
<td>92%</td>
<td>43%</td>
<td>53%</td>
<td>56%</td>
</tr>
<tr>
<td>Y147F</td>
<td>70%</td>
<td>42%</td>
<td>41%</td>
<td>28%</td>
</tr>
<tr>
<td>Y147M</td>
<td>97%</td>
<td>62%</td>
<td>59%</td>
<td>56%</td>
</tr>
<tr>
<td>Y147H</td>
<td>66%</td>
<td>9%</td>
<td>19%</td>
<td>2%</td>
</tr>
<tr>
<td>Y147N</td>
<td>41%</td>
<td>4%</td>
<td>47%</td>
<td>11%</td>
</tr>
<tr>
<td>Y147R</td>
<td>24%</td>
<td>4%</td>
<td>34%</td>
<td>17%</td>
</tr>
<tr>
<td>Y147C</td>
<td>25%</td>
<td>3%</td>
<td>28%</td>
<td>22%</td>
</tr>
<tr>
<td>Y147G</td>
<td>24%</td>
<td>3%</td>
<td>25%</td>
<td>3%</td>
</tr>
<tr>
<td>Y147T</td>
<td>23%</td>
<td>5%</td>
<td>28%</td>
<td>6%</td>
</tr>
<tr>
<td>Y147E</td>
<td>22%</td>
<td>2%</td>
<td>38%</td>
<td>17%</td>
</tr>
<tr>
<td>Y147S</td>
<td>14%</td>
<td>8%</td>
<td>31%</td>
<td>17%</td>
</tr>
<tr>
<td>Y147D</td>
<td>11%</td>
<td>1%</td>
<td>25%</td>
<td>6%</td>
</tr>
<tr>
<td>Y147V</td>
<td>8%</td>
<td>1%</td>
<td>16%</td>
<td>3%</td>
</tr>
<tr>
<td>Y147P</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>
catalytic and structural information about both the family 7 and the family 45 cellulases, but still it is not possible to explain the different pH activity profiles of the cellulases. The highly conserved amino acids directly involved in binding of the substrate and for catalysis are conserved but do not explain the variation in pH optima.


Received 4 December 1997

Structural studies on cellulases
G. J. Davies
Department of Chemistry, University of York, Heslington, York YO1 5DD, U.K.

Introduction
The enzymatic breakdown of the β-1,4 links of cellulose is performed by an ensemble of enzymes, the cellulases, which perform general acid/base hydrolysis of the glycosidic bonds. These enzymes have received extensive sequence, structural and biochemical characterization in recent years. At the three-dimensional level, the structural analysis has undoubtedly benefited by recent technical advances in protein crystallography. In particular, the development of image-plate detectors, high-intensity synchrotron X-ray sources and the introduction of cryocrystallographic techniques means that structural analysis can now be completed in a fraction of the time previously required. Experiments with ligand complexes of enzymes become extremely rapid and even a series of atomic-resolution analyses becomes tractable on a reasonable timescale.

Cellulases are a subset of enzymes which hydrolyse the glycosidic bond and collectively are termed the glycoside hydrolases. The first enzyme structure ever solved was a glycoside hydrolase, hen egg-white lysozyme (HEWL) [1]. More recently there has been an explosion in the determination of glycoside hydrolase structures [2,3]. Representatives for many of the sequence families of these enzymes (described below) have been described at the three-dimensional level. The synergistic combination of X-ray structural analysis together with increasingly elegant meth-