# Characterization of a *Thermomonospora fusca* Exocellulase[†,‡]

Sheng Zhang, Guifang Lao, and David B. Wilson*

*Section of Biochemistry, Molecular, and Cell Biology, Cornell University, Ithaca, New York 14853*

*Received August 4, 1994; Revised Manuscript Received November 28, 1994*[⊗]

ABSTRACT: The exocellulase E3 gene was cloned on a 7.1 kb *Not*I fragment from *Thermomonospora fusca* genomic DNA into *Escherichia coli* and expressed in *Streptomyces lividans.* The E3 gene was sequenced and encoded a 596 residue peptide. The molecular masses of the native and cloned E3s were determined by mass spectrometry, and the value for *E. coli* E3, 59 797 Da, agreed well with that predicted from the DNA sequence, 59 646 Da. The value of 61 200 Da for *T. fusca* E3 is consistent with E3 being a glycoprotein. E3 is thermostable, retaining full activity after 16 h at 55 °C. It also has a broad pH optimum around 7−8, retaining 90% of its maximal activity between pH 6 and 10. The cloned E3s were identical to the native enzyme in their activity, cellulose binding, and thermostability. Papain digestion produced a 45.7 kDa catalytic domain with 77% of the native activity on amorphous cellulose and 33% on crystalline cellulose. E3 belongs to cellulase family B and retains the residues that have been identified to be crucial for catalytic activity in *Trichoderma reesei* cellobiohydrolase II and *T. fusca* E2. The E3 gene contains a 14 bp inverted repeat regulatory sequence 212 bp before the translational start codon instead of the 30−70 bp found for the other *T. fusca* cellulase genes. An additional copy of this sequence with one base changed is 314 bp before the translational start codon. The transcriptional start site of the E3 gene was shown to be between these two inverted repeats.

Cellulases degrade insoluble cellulose into glucose oligomers, and they can be classified into two broad groups: endocellulases (EC 3.2.1.4) and exocellulases (EC 3.2.1.91). One reason cellulases have been actively studied is their potential application for the conversion of biomass to ethanol (Dale, 1987; Lynd et al., 1991). They are also mechanistically interesting because the maximal rate of cellulose degradation requires the action of several different enzymes whose activities complement one another (Klyosov, 1990).

We have been studying the cellulases of a thermophilic actinomycete, *Thermomonospora fusca* (Wilson, 1992). At least six structurally and functionally distinct cellulases, E1−E6, which show synergism with each other have been purified and characterized from the culture supernatant of a protease-negative mutant of *T. fusca* (Irwin et al., 1993). The structural genes for four of these, E1, E2, E4, and E5, have been cloned (Collmer & Wilson, 1983; Ghangas & Wilson, 1988; Hu & Wilson, 1988), sequenced, and expressed (Lao et al., 1991; Jung et al., 1993). They belong to three different cellulase families on the basis of the amino acid sequences of their catalytic domains (Gilkes et al., 1991b).

E3, a typical exocellulase, is very important in achieving maximum activity of synergistic mixtures although its activity by itself is quite low (Irwin et al., 1993). It resembles the fungal exocellulase cellobiohydrolase II (CBHII)[1] from *Trichoderma reesei* in its function (Irwin et al., 1993) but has significantly higher thermostability and a broader pH optimum. In this paper, we report the cloning, sequencing, and expression of the E3 gene from *T. fusca* as well as the characterization of E3 and its catalytic domain.

## MATERIALS AND METHODS

*Bacterial Strains and Plasmids.* The bacterial strains and plasmids used in this study are listed in Table 1. Plasmids pSZ3 to pSZ12 were constructed as described below. *E. coli* DH5α (Hanahan, 1983) was used for cloning, subcloning, and plasmid isolation. *S. lividans* TKM31, used as the host cell for expression, is a protease-negative strain we isolated from *S. lividans* TK24 which was obtained from D. A. Hopwood, John Innes Institute, Norwich, England.

*Growth of Organisms.* Unless indicated, *T. fusca* was grown as described previously (Calza et al., 1985). *E. coli* strains containing recombinant plasmids were grown in Luria broth and plated on Luria agar plates containing 100 μg/mL ampicillin. Cultures for the purification of E3 were grown in superbroth medium (Sambrook et al., 1989). Media for sporulation of *S. lividans* as well as the preparation and regeneration of protoplasts were as described (Hopwood et al., 1985). *S. lividans* cultures were grown on tryptone soya broth as reported previously (Ghangas & Wilson, 1988). Thiostrepton was used at a concentration of 10 μg/mL in liquid media and 50 μg/mL in plates.

*N-Terminal Amino Acid Sequence of an E3 Cyanogen Bromide Peptide.* E3 was purified from the culture supernatant of *T. fusca* as described by Irwin (Irwin et al., 1993). Cyanogen bromide cleavage of E3 (3 mg) was performed by the method of Villa (Villa et al., 1989). The E3 fragments were separated by SDS−PAGE (Laemmli, 1970) and electrophoretically blotted onto Immobilon-P membranes (PVDF

[†] Supported by the Cooperative State Research Service, U.S. Department of Agriculture, under Agreement 91-37206-6737.

[‡] The DNA sequence data reported in this paper were submitted to Genebank under Accession Number U18978.

[⊗] Abstract published in *Advance ACS Abstracts,* February 15, 1995.

[1] Abbreviations: CBHII, cellobiohydrolase II; Tsr, thiostrepton; CMC, (carboxymethyl)cellulose; TSB, tryptone soya broth; PMSF, phenylmethanesulfonyl fluoride; GL, glucanolactone; *S*E3, purified exocellulase E3 from *Streptomyces lividans; E*E3, purified exocellulase E3 from *E. coli*; *T*E3, purified exocellulase E3 from *Thermomonospora fusca*; E3cd, catalytic domain of exocellulase E3; CBHI, cellobiohydrolase I; CBHIIcd, catalytic domain of cellobiohydrolase II; CBD, cellulose binding domain; CNBr, cyanogen bromide.

0006-2960/95/0434-3386$09.00/0 

Table 1: Host Strains and Plasmids Used in This Study

| plasmid | host | description[a] | source or reference |
|---|---|---|---|
| pUC18 | *E. coli* DH5α | 2.69 kb, ampicillin resistance | *b* |
| pGG82 | *S. lividans* TK24 | 6.2 kb, derivative of pIJ702 with Mel− phenotype | *c* |
| pBluescript SK+ | *E. coli* DH5α | 2.96 kb, ampicillin resistance | Stratagene |
| pSZ3 | *E. coli* DH5α | 7.1 kb E3 gene in pBluescript SK+ (− orientation to *Lac*) | this paper |
| pSZ4 | *E. coli* DH5α | 7.1 kb E3 gene in pBluescript SK+ (+ orientation to *Lac*) | this paper |
| pSZ5 | *E. coli* DH5α | 3.0 kb E3 gene in pUC18 (− orientation to *Lac*) | this paper |
| pSZ6 | *E. coli* DH5α | 3.0 kb E3 gene in pUC18 (+ orientation to *Lac*) | this paper |
| pSZ7 | *E. coli* DH5α<br>*S. lividans* TKM31 | 5.7 kb pSZ6 in pGG82 (+ orientation to *Tsr*) | this paper |
| pSZ7A | *S. lividans* TKM31 | 8.6 kb, derivative of pSZ7 without pUC18 (+ orientation to *Tsr*) | this paper |
| pSZ8 | *E. coli* DH5α<br>*S. lividans* TKM31 | 5.7 kb pSZ5 in pGG82 (− orientation to *Tsr*) | this paper |
| pSZ9 | *E. coli* DH5α | 4.1 kb E3 gene in pUC18 (− orientation to *Lac*) | this paper |
| pSZ10 | *E. coli* DH5α | 4.1 kb E3 gene in pUC18 (+ orientation to *Lac*) | this paper |
| pSZ11 | *E. coli* DH5α<br>*S. lividans* TKM31 | 6.8 Kb pSZ10 in pGG82 (+ orientation to *Tsr*) | this paper |
| pSZ12 | *E. coli* DH5α<br>*S. lividans* TKM31 | 6.8 kb pSZ9 in pGG82 (− orientation to *Tsr*) | this paper |

[a] E3 gene aligned opposed to the *Lac* or *Tsr* promoter. +E3 gene aligned with the *Lac* or *Tsr* promoter. [b] Bethesda Research Laboratory (1986). [c] Ghangas et al. (1989).

membranes from Millipore) as described (Matsudaira, 1987). A 12 kDa band (E3p12) was cut out, and its N-terminal sequence was determined using an Applied Biosystems 470A protein sequenator.

*Cloning Procedures and Recombinant DNA Techniques. (A) Preparation of E3 Probes.* The N-terminal amino acid sequence of E3 was reported earlier (Wilson, 1988). A degenerate set of 20-mer oligonucleotides [TGGGGG(C)-ACG(C)GGG(C)TTC(T)ACC(dI)GC] encoding the N-terminal region from Trp to Thr of E3 was synthesized and labeled with digoxingenin−dUTP by the 3′-tailing reaction with the Genius DNA Labeling and Detection kit (Boehringer Mannhein).

Since this probe hybridized to several bands, a 20-mer degenerate oligonucleotide, ACG(A)TAC(G)CCC(G)CCG-(A)TTC(T)TGC(T)TT, complementary to the strand encoding the first six N-terminal amino acids of E3p12 was prepared and labeled as described for the E3 probe.

*(B) Cloning of the E3 Gene. T. fusca* chromosomal DNA was isolated as reported previously (Collmer & Wilson, 1983). Mini- and mid-scale preparations of *E. coli* plasmid DNA were performed by using the INSTA-PREP kit and the plasmid SELECT-100 column kit (5 Primer and 3 Primer Inc). Transformation of *E. coli* was carried out by electroporation and by the TSS method (Dower et al., 1988; Chung et al., 1989).

Genomic Southern hybridization was carried out using total *T. fusca* DNA digested completely with *Not*I and electrophoresed on a 0.7% agarose gel as described (Jung et al, 1993). One positive band, located at 7.1 kb, was found on the filter hybridized to the E3p12 probe, and it also hybridized to the E3 probe.

Fragments around 7.0 kb were cut out from a 0.8% low melting agarose gel, run on a *Not*I total digest of *T. fusca* DNA, purified by β-agarase treatment (Burmeister & Lehrach, 1989), and then ligated to pBluescript SK+ (Stratagene, La Jolla, CA) that had been digested with *Not*I and dephosphorylated. The ligation mixtures were used to transform *E. coli* DH5α and plated on LB Amp plates containing X-gal and IPTG (Sambrook et al., 1989). Transformants were screened by *Not*I digestion of plasmid DNA followed by Southern hybridization to both probes and by a CMC overlay assay at 50 °C for 24 h (Teather & Wood, 1982).

Eleven plasmids were constructed: pSZ3 and pSZ4 have the 7.1 kb *Not*I−*Not*I fragment cloned into pBluescript in opposite orientations; pSZ5 and pSZ6 contain a 3 kb *Pst*I−*Pst*I fragment from plasmid pSZ3 in opposite directions subcloned into pUC18.

The shuttle plasmids pSZ7 and pSZ8 were constructed by inserting, respectively, pSZ6 and pSZ5 into pGG82. pSZ6, pSZ5, and pGG82 were digested with *Hin*dIII and *Sph*I. The large fragments from pSZ6, pSZ5, and pGG82 were isolated on a low melting agarose gel, ligated, and transformed into *E. coli* DH5α. The desired transformants were identified by restriction mapping of plasmid DNAs and by the CMC overlay assay. Plasmid pSZ7A is a derivative of pSZ7 that was constructed by cutting out pUC18 with *Eco*RI. The 3 kb fragment in pSZ5/6 lacked the potential E3 promoter region, so pSZ9 and pSZ10 which contained the whole E3 gene were constructed by inserting a 4.1 kb fragment from a *Pst*I partial digest of pSZ3 into pUC18 as described for pSZ5/6. The shuttle plasmids pSZ11 and pSZ12 were constructed by inserting, respectively, pSZ10 and pSZ9 into pGG82 as described for the construction of pSZ7.

To express E3 in *S. lividans*, pSZ7, pSZ7A, pSZ8, pSZ11, and pSZ12 were transformed into *S. lividans* protoplasts and plated on $R_2$YE plates (Hopwood et al., 1985). Figure 1A shows the cloning strategy and the plasmids used for expression.

*DNA Sequencing.* Double-stranded DNA from pSZ6 and pSZ4 was used for sequencing the E3 structural gene and its 3′ and 5′ flanking regions. The sequences of both strands of the E3 gene were determined by the dideoxy-chain termination method (Sanger et al., 1977) with a Sequenase kit from U.S. Biochemical Corp. The two degenerate 20mer oligonucleotides and the universal primers for pUC/M13 sequencing were used to determine the initial sequences within the inserts, and then specific primers for regions within the inserts were designed and synthesized by the Oligonucleotide Synthesis Facility, Cornell University. The DNA sequencing procedure and the programs used to analyze DNA sequences were carried out as described (Jung et al., 1993).

*Primer Extension Analysis.* Total RNA was purified (Summers, 1970) from an *E. coli* strain containing pSZ4. The oligonucleotide 5′ GCATCCACGAACGTCTGTTCGTG-GCACGAACTTTACTCAT3′ complementary to a region coding for the signal peptide of E3 was synthesized and labeled at its 5′ terminus with [$\gamma$-$^{32}$P]ATP and polynucleotide kinase. Northern hybridization was then carried out as described previously (Lin & Wilson, 1988) to determine the size of E3 mRNA, followed by primer extension experiments (Kinston, 1989). Total RNA (50 $\mu$g) was hybridized to the $^{32}$P-labeled oligonucleotide. AMV reverse transcriptase was added, and the labeled transcripts from the oligonucleotide extension and regular DNA sequencing mixtures using the same oligonucleotide with pSZ4 plasmid DNA were electrophoresed on a denaturing 6% polyacrylmide sequencing gel that was autoradiographed.

*Purification of E3 from S. lividans.* A 3-day culture of *S. lividans* pSZ7 (25 mL) in TSB was used to inoculate 250 mL of the same medium. After growth for 48 h at 30 °C, the entire culture was added to 10 L of the same medium. Mycelia were harvested after 72 h of fermentation (agitation, 200 rpm; air flow, 1 volume of air per volume of medium per minute; temperature, 30 °C; pH, initially at 7.1) by cross-flow filtration with a Millipore Pelicon cassette equipped with 0.45 $\mu$m membranes.

All purification procedures were carried out at 4 °C. Phenylmethanesulfonyl fluoride (PMSF) and ammonium sulfate were added to mycelia-free supernatant at 0.1 mM and 1 M final concentrations, respectively. The supernatant was loaded onto a CL-4B phenyl-Sepharose column (10 × 14 cm) which was equilibrated with 0.6 M $(NH_4)_2SO_4$, 0.01 M NaCl, and 0.005 M $KP_i$, pH 6.0. The column was washed with 2 volumes of equilibration buffer, followed by 2 volumes of 0.3 M $(NH_4)_2SO_4$, 0.01 M NaCl, and 0.005 M $KP_i$, pH 6.0, and then the protein was eluted with 0.005 M $KP_i$, pH 6. The fractions containing activity were combined and adjusted to 0.25 M $(NH_4)_2SO_4$/1 mM glucanolactone (GL) by adding solid $(NH_4)_2SO_4$ and GL and applied to a *p*-nitrobenzyl 1-thio-$\beta$-D-cellobioside affinity column (Van Tilbeurgh et al., 1984) that was equilibrated with 0.25 M $(NH_4)_2SO_4$, 1 mM GL, and 0.005 M $KP_i$, pH 6.0. After the column was loaded, it was washed with 2 volumes of 0.1 M $(NH_4)_2SO_4$, 1 mM GL, and 0.005 M $KP_i$, pH 6.0, and then with 2 volumes of 0.1 M NaOAc buffer, pH 5.5, + 1 mM GL. E3 was eluted by the addition of 0.1 M lactose to the wash buffer. The appropriate fractions were finally applied to a Q-Sepharose column equilibrated with 0.1 M NaOAc buffer, pH 5.5. After the column was washed with 0.2 M NaCl/0.02 M Bis-Tris, pH 5.1, a linear gradient from 0.2 to 0.5 M NaCl was used to elute *S. lividans* E3 (*S*E3) that was at least 95% pure by SDS−PAGE.

*Purification of E3 from E. coli.* An overnight superbroth culture (10 mL) of *E. coli* pSZ6 strain was inoculated into 1 L of the same medium. The culture was grown for 22 h with rotary shaking at 37 °C and centrifuged at 4500*g* for 15 min; the pellet was resuspended in 50 mL of 0.05 M NaOAc, pH 5.5, + 1 mM PMSF, French-pressed at 10 000 lb/in.$^2$, and centrifuged at 14000*g* for 30 min. The lysate was adjusted to 0.25 M $(NH_4)_2SO_4$ + 1 mM GL and loaded onto a cellobioside affinity column (2.5 × 10 cm) that was previously equilibrated with the same buffer. The column was washed with 2 volumes of equilibration buffer and 2 volumes of 0.05 M NaOAc, pH 5.5, + 1 mM GL, and then 0.02 M cellobiose in wash buffer was used to elute E3. Further purification was performed on the Q-Sepharose column as described above. *E. coli* E3 (*E*E3) was 95% pure by SDS−PAGE.

*Limited Proteolysis of E3 by Papain and Isolation of E3cd.* To investigate the possible domain structure of E3 and to compare the differences in the proteolytic resistance of *T. fusca* E3 (*T*E3) and cloned E3 (*S*E3 and *E*E3), purified E3 was partially digested with papain as follows: 15 $\mu$L of 1.0 mg/mL papain (Sigma) solution in 0.05 M $NH_4OAc$ buffer, 5 mM L-cysteine, and 2 mM EDTA, pH 6.5, was added to 100 $\mu$g of purified E3 from each source in 85 $\mu$L of 0.05 M $NH_4OAc$, pH 6.5. The mixtures were incubated at 37 °C, and aliquots (2−10 $\mu$L) were removed at 0.5, 1, 3, 6, and 24 h for further analysis by both SDS−PAGE and native PAGE followed by a CMC overlay. To isolate the catalytic domain of E3 (E3cd), 15 mg of *T*E3 was partially digested by papain for 1 h and chromatographed by gel filtration on an Ultrogel AcA54 column (2.6 × 100 cm) with 0.05 M NaOAc buffer, pH 5.5. The eluates were analyzed by a SDS−gel, and the E3cd was homogenous.

*Preparation of Trichoderma reesei CBHI and CBHII. T. reesei* CBHI and CBHII were purified as described (Irwin et al., 1993).

*Activity, Binding Assays, and Protein Determination.* All activity assays were run in triplicate. CMCase assays were run at 50 °C for 16 h with 1% Sigma low-viscosity CMC or 0.25% swollen cellulose as substrate while filter paper activities were determined as described (Irwin et al., 1993) and synergistic filter paper reactions were set up to contain 20% E5 and equal molar amounts of *T. ressei* CBHI, E3, or E3cd. All proteins were measured by their absorbance at 280 nm using extinction coefficient values calculated from their predicted amino acid compositions.

Binding of E3 and E3cd to cellulose was determined by adding 266 $\mu$g of E3 and E3cd to 0, 5, 10, 20, 50, and 100 mg of Avicel PH102(FMC) in 1 mL of 0.05 M NaOAc buffer, pH 5.5. The samples were incubated at 50 °C for 1 h with end-over-end rotation. After centrifugation, the E3 left in the supernatant was measured by its absorbance at 280 nm.

A LASER MAT matrix-assisted laser desorption mass spectrometer was used to determine the molecular mass of E3 and E3cd isolated from three different sources using either external calibration or ribonuclease as an internal standard with the laser power attenuated to 40−95% and the laser aim set at 1−4.

*N-Terminal Sequence and Amino Acid Analysis of Purified E3 and E3cd.* After SDS−PAGE electrophoresis, the purified E3 and E3cd were electroblotted onto PVDF membranes as described above, and the corresponding bands were cut out. The N-terminal amino acid sequences were determined by Edman degradation with an Applied Biosystem 470A automated gas-phase protein sequencer.

Amino acid compositions were determined by the Oligonucleotide Synthesis Facility, Cornell University. Samples were hydrolyzed in 6 N HCl for 105 min at 150 °C followed by column fractionation on a Water Pico Tag analyzer.

*Western Blots.* E3 produced by the different strains in supernatants and cell extracts was identified by Western blots (Towbin et al., 1979). The proteins were separated by 12% SDS−PAGE, and electrophoretically transfered to a nitrocellulose membrane and reacted with rabbit polyclonal E3 antibody. The E3 bands were detected with goat anti-rabbit immunoglobulin G conjugated to alkaline phosphatase and

visualized with nitroblue tetrazolium chloride and 5-bromo-4-chloro-3-indolyl phosphate.

*Glycosylation Detection.* Protein glycosylation was detected with the DIG glycan detection kit (Boehringer Mannheim) using 2−10 μg of protein.

## RESULTS

About 150 transformants from the *Not*I library were screened with the E3p12 probe, and 4 were positive. The positive colonies were tested by CMC overlay assay and by Western blotting using E3 antibodies, and all four were positive by both assays. Plasmid DNA was prepared from all four transformants, and restriction mapping showed that all of them contained the same 7.1 kb *Not*I fragment (Figure 1B) but pSZ4 had the E3 gene in the same orientation as the *lac* promoter while pSZ3 had the E3 gene in the opposite orientation

pSZ3 was digested with *Pst*I, and a 3.0 kb fragment was subcloned into pUC18. Two plasmids, pSZ5 and pSZ6, containing the insert in different directions were identified. Both CMC overlay and Western blotting showed that this fragment expressed E3.

*Expression of E3 in S. lividans.* E3 was expressed and secreted by all the *S. lividans* transformants. In order to compare the effects of gene orientation and the E3 promoter region on the level of expression in *S. lividans*, the yield of E3 produced by each type of transformant was determined based on activity assays on swollen cellulose and filter paper using a known amount of purified E3 in the supernatant of a control *S. lividans* culture as a standard. The calculated yields of E3 for each plasmid are 18 mg/L of TSB culture for pSZ7, 38 mg/L for pSZ7A, <4.0 mg/L for pSZ8, 36 mg/L for pSZ11, and 42 mg/L for pSZ12. From the E3 activity of an *E. coli* extract, E3 was expressed at a 5−8-fold higher level in *S. lividans* supernatant than in *E. coli* cells.

*Nucleotide Sequence Analysis.* The 3.0 kb insert in pSZ6 was used for sequencing the E3 structural gene and its 3′ flanking region while the 7.1 kb insert in pSZ4 was used to sequence its 5′ flanking region. Both strands were sequenced as shown in Figure 1B, and the sequence is shown in Figure 2. The cloned DNA had a G+C content of 66%, which agrees well with the 65% G+C content of *T. fusca* DNA reported previously (Lao et al., 1991) and the 67% G+C content of *T. curvata* DNA (Petricek et al., 1989). The correct reading frame was determined by computer analysis based on the high G+C content of the third position of codons. A reading frame from nucleotides 475 to 2262 (Figure 2) encodes a 596 amino acid protein that corresponds to the E3 precursor and has a G+C content of 91% in the third positions of the codons.

The open reading frame encoding the E3 gene probably begins with the ATG at nucleotide 475 since a potential ribosome binding site (AAGGA) is present 10 bases upstream.

A single transcriptional start site was determined by primer extension to be at nucleotide 162, 314 bases upstream of the translational initiation codon. The size of E3 mRNA estimated by Northern hybridization is 2.0−2.1 kb, which is in good agreement with the deduced size of the reading frame and the 5′ untranslated region (2.1kb).

A 14 bp inverted repeat with the sequence 5′TGGGAGCGCTCCCA3′ was located 212 bases before the translational start codon rather than the 34−62 bases found for four other *T. fusca* cellulase genes (Jung et al., 1993; Lao et al., 1991); a closely related sequence was present 314 bases before the translational start codon.

E3 has a signal peptide of 38 amino acids preceding its native N-terminal sequence, which resembles other actinomycete signal sequences in its size, composition, and cleavage site (Hutter & Eckhardt, 1988).

A sequence similar to a Rho-independent terminator in *E. coli* (Rosenberg & Court, 1979) is downstream of the E3 structural gene, which contains a 14 base palindrome followed by four Ts.

*Protein Size and Composition.* The molecular mass of *T. fusca* E3 (*T*E3, 65 000 Da) was reported previously from SDS−PAGE (Irwin et al., 1993; Wilson, 1988). All three forms of E3 display nearly identical electrophoretic mobilities on an SDS gel (Figure 3A). On a native polyacrylamide gel, *E*E3 was found to move a little faster than *T*E3 and *S*E3 (Figure 3B). The N-terminal sequences of the cloned E3s and of *T*E3 were identically showing that all three organisms use the same site for signal peptide cleavage. However, *S*E3 alone had AAPANAAGCS as a second N-terminal sequence, showing there is an additional cleavage site, six amino acids before the regular site. The second site appears to be used 70% of the time.

Amino acid analyses of the three different species were nearly identical (average error 5% per residue with a range of 2−12%) and agreed well with the predicted composition (average error 7% with a range of 0−15%). The predicted molecular mass of the mature protein is 59 646 Da, which is lower than that estimated from SDS−PAGE (65 000 Da). This value is consistent with that of *E*E3 (59 797 Da) as determined by mass spectrometery while that of *T*E3 (61 200 Da) is larger, as expected for a glycoprotein, and that for *S*E3 (61 169 Da) is about 900 Da larger than the value calculated from its N-terminus, which is constant with it also being a glycoprotein. The molecular mass of the peptides derived from cyanogen bromide cleavage of *T*E3 were also measured by mass spectrometry. *T*E3 cleaved by CNBr released six peptides with molecular masses of 21 745, 7310, 4374, 13 800, 9957, and 4748 Da. These values correlate well with the values of 20 084, 6671, 4288, 14 100, 9865, and 4728 Da predicted from the positions of the methionine residues in the sequence. The only significant differences between the predicted and measured values are in the first two peptides which are 10% higher than predicted. The first peptide contains the cellulose binding domain and the linker peptide which contains the carbohydrate residues. This result is similar to that reported for other glycosylated cellulases (Van Tilbeurgh et al., 1986; Tomme et al., 1988; Gilkes et al., 1988).

*Isolation of an E3 Catalytic Domain.* Unlike E2 and E5 (Ghangas & Wilson, 1987, 1988), cloned E3 appears to be stable in vivo. Even after 3−5 days of growth in TES-Hag medium (Hopwood et al., 1985), secreted *S*E3 remained intact. Thus, partial digestion of E3 by papain was used to produce an E3 catalytic domain. As shown in Figure 4A, intact E3 was completely converted to a 46 kDa fragment (E3cd) by a 60 min digestion under the given conditions (see Materials and Methods). No further degradation appeared even after 24 h. The only bands visualized by Coomassie blue staining were intact E3 and E3cd. It is likely that the binding domain and linker region have been degraded. The molecular mass of E3cd on SDS−PAGE was not influenced by reduction, suggesting that no additional

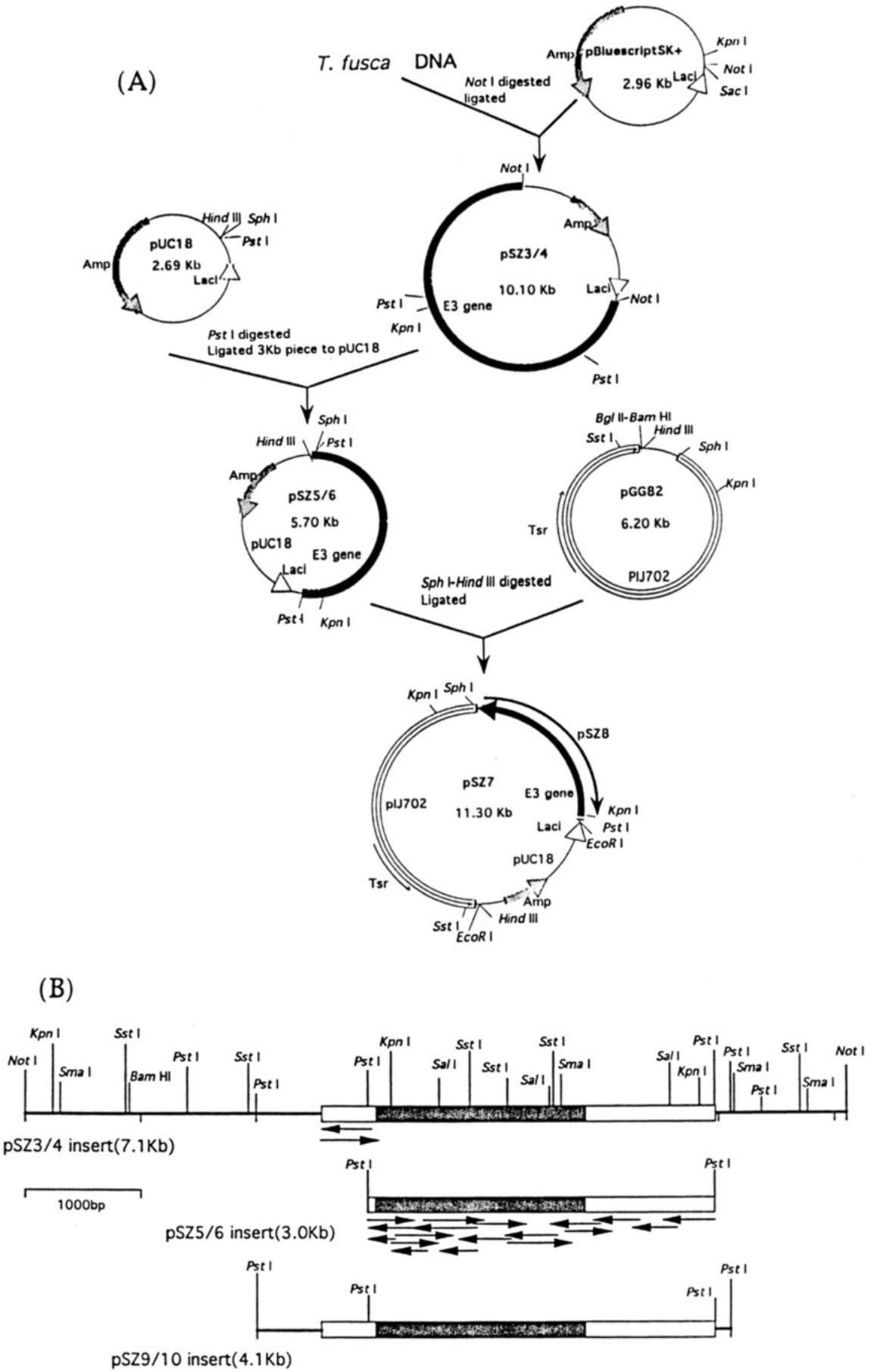


FIGURE 1: (A) Cloning and subcloning of an exocellulase (E3) gene from *Thermomonospora fusca* into *E. coli* and *Streptomyces lividans*. Thin black lines denote pUC18/pBluescript SK+ sequences, the open and lined curve indicates the pIJ702 sequence, and thick black curves represent *T. fusca* DNA. (B) A physical map of *Not*I and *Pst*I fragments from *T. fusca*. The arrows represent the regions and directions of sequencing by universal and synthetic primers. The bar indicates sequenced *T. fusca* DNA, stippled boxes represent the E3 structural gene, and the thick solid bar is unsequenced *T. fusca* DNA.

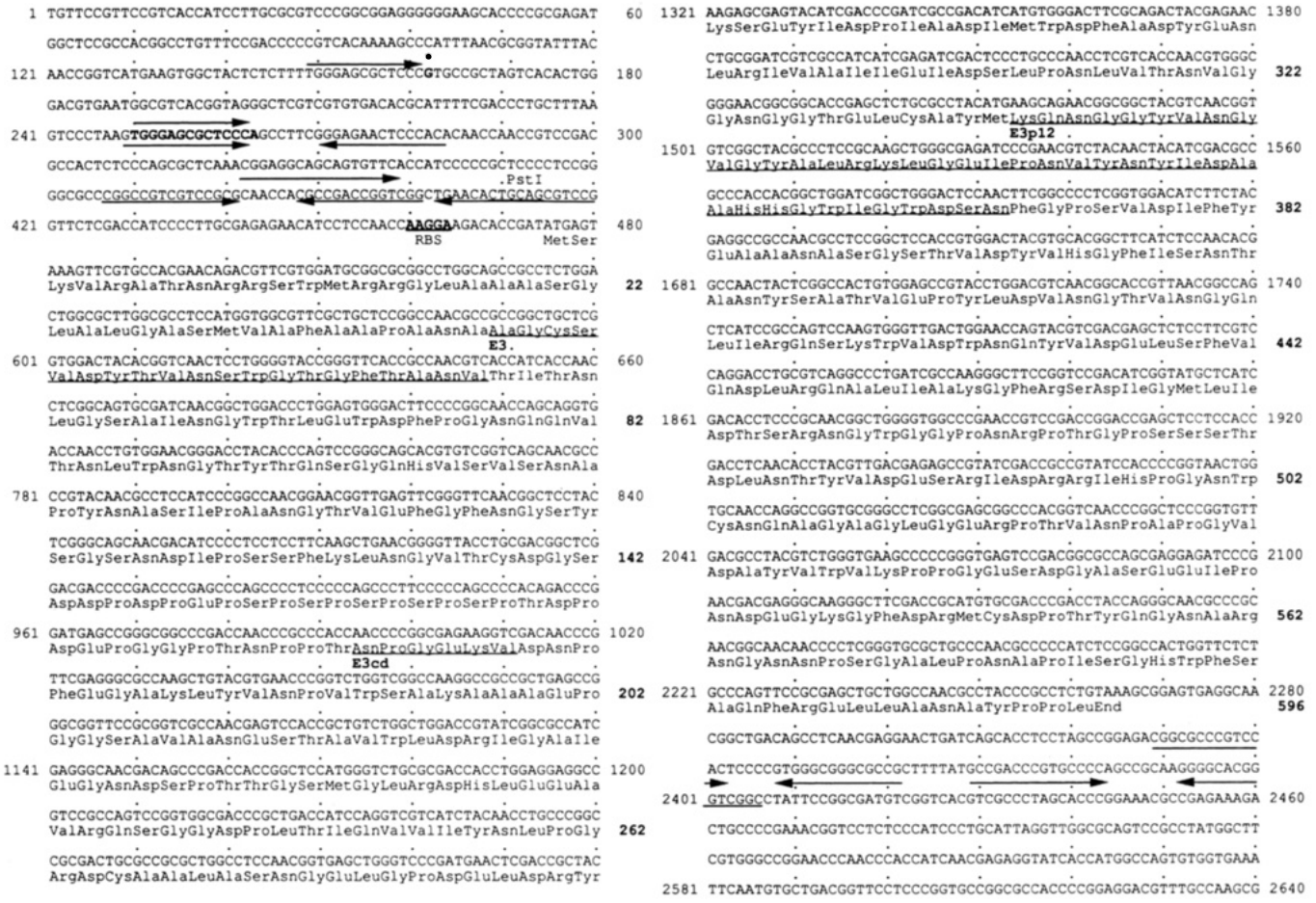


FIGURE 2: DNA sequence and deduced amino acid sequence of the E3 gene. The numbers on both sides indicate nucleotide positions, and the boldface numbers on the right are the amino acid position. An asterisk marks the transcriptional start site. The thick horizontal arrows above the sequences mark direct repeats, and facing arrows beneath the sequences indicate inverted repeat sequences. The 14 bp inverted repeat is highlighted. The N-terminals of E3, E3cd, and E3p12 (CNBr cleavage product) are underlined and noted. RBS, ribsome binding site. Restriction enzyme sites are noted by the name above the sites.

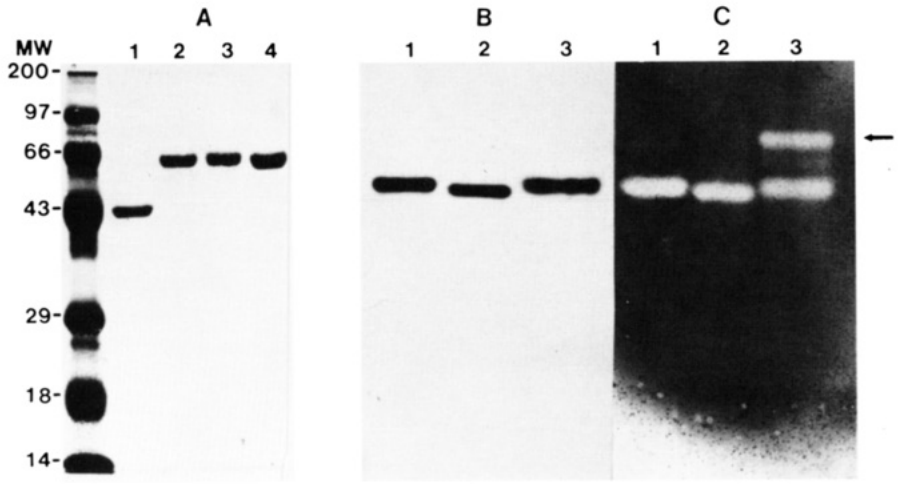


FIGURE 3: (A) Electrophoresis of purified *T. fusca* E3, E3cd, and the cloned enzymes from *E. coli* and *S. lividans*. Estimation of the molecular mass of E3 determined on a 12% SDS gel. Two microgram of each sample was loaded. Lanes: 1, E3cd; 2, *T*E3; 3, *S*E3; 4, *E*E3. The molecular mass markers are indicated in kilodaltons at the left. (B, C) Migration of E3 on an 8.5% native polyacryalmide gel. Stained with Coomassie brilliant blue R250 (B) and with Congo Red to detect CMCase activity (C). Lane 1, 10 $\mu$g of *T*E3; lane 2, 10 $\mu$g of *E*E3; lane 3, 10 $\mu$g of *S*E3. The top band indicated by an arrow in lane 3 of panel C was a contaminating CMCase from *S. lividans*.

cleavage site existed in E3cd. A CMC overlay of a native gel showed that E3cd possessed similar CMCase activity to E3 (data not shown).

The exact molecular mass of the E3cd produced from each form of E3 was determined by mass spectrometry. All three forms gave a value of 46 kDa (*T*E3cd, 46 092 Da; *S*E3cd, 46 013 Da; and *E*E3cd, 46 067 Da), which is very close to the predicted molecular mass of 45 707 Da. The fact that all three forms of E3cd have the same molecular mass indicates that all of the sugar in *T*E3 is present in the binding domain and linker region, which is consistent with the results of the determination of the molecular mass of the CNBr peptides and the glycosylation assay of *T*E3cd as shown in Figure 4B,C.

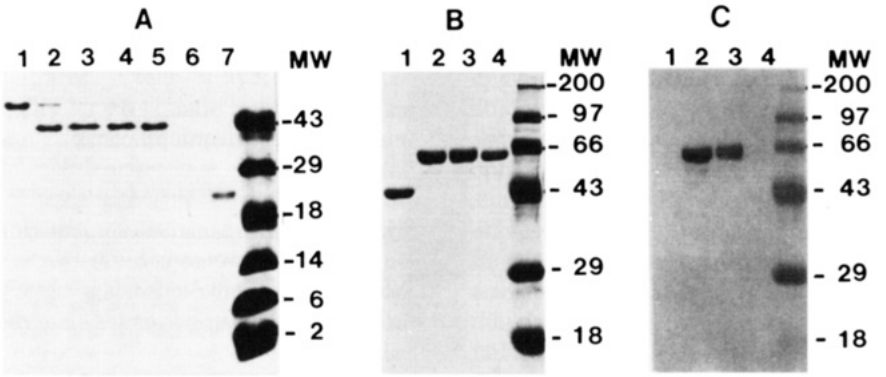


FIGURE 4: (A) Partial papain digest of E3 run on 16% SDS−PAGE. Lanes 1−5, after 0, 0.5, 1, 6, and 24 h; lanes 6 and 7, 20 and 200 ng of papain alone. (B, C) Glycosylation analysis of *T*E3 and *S*E3 on a 12% SDS gel. Two micrograms of sample was loaded on a SDS gel after being labeled with digoxigenin−3-*O*-ϵ-aminocaproic acid hydrazide hydrochloride, stained with Coomassie blue (B), or electroblotted onto a NC membrane and stained with the DIG glycan detection kit (C). Lanes 1−4, *T*E3cd, *T*E3, *S*E3, and *E*E3. The molecular mass markers are labeled in kilodaltons at the right.

Table 2: Activity Assay of Native E3, Cloned E3s and E3cd

| | activity [$\mu$mol of CB min$^{-1}$ ($\mu$mol of enzyme)$^{-1}$][b] | | |
|---|---|---|---|
| enzyme[a] | CMC | filter paper | swollen cellulose |
| *T*E3 | 0.62[d] | 0.153[d] | 1.40 |
| *S*E3 | 6.53[c] | 0.373[c,d] | 6.25[c] |
| *E*E3 | 0.65[d] | 0.157[d] | 1.52 |
| E3cd | 0.48[d] | 0.050[d] | 1.09 |
| E5 | 2840.0 | 0.682[d] | 90.0 |
| CBHI | 2.1 | 0.484[d] | 9.3 |
| *T*E3+E5 | ND[e] | 2.61 | ND |
| *S*E3+E5 | ND | 2.89 | ND |
| *E*E3+E5 | ND | 2.85 | ND |
| E3cd+E5 | ND | 0.622[d] | ND |
| *T*E3+E5cd | ND | 2.15 | ND |
| E3cd+E5cd | ND | 0.632[d] | ND |
| *T*E3+E5+CBHI | ND | 7.46 | ND |
| *S*E3+E5+CBHI | ND | 6.52 | ND |
| *E*E3+E5+CBHI | ND | 7.56 | ND |
| E3cd+E5+CBHI | ND | 3.37 | ND |
| *T*E3+E5cd+CBHI | ND | 6.53 | ND |
| E3cd+E5cd+CBHI | ND | 3.87 | ND |

[a] Molar ratios of mixture components were 4:1 for the E3:E5 group and 2:1:2 for the E3:E5:CBHI group. [b] Extinction coefficients for E3 (115 150/molar) and E3cd (87 150/molar) were determined from the predicted sequence. [c] Contaminating CMCase activity as determined by a CMC overlay of a native gel. [d] Target percentage (5.2%) digestion could not be achieved; activities were calculated using digestion achieved by 0.6 nmol of enzyme in 16 h. [e] ND, not determined.

*Activity and Binding Properties of the Different E3s and E3cd.* CMC, swollen cellulose, and filter paper were used as substrates for determining the specific activities of all three forms of E3 and E3cd. Although E3 by itself has quite low activity on all three substrates, it was previously shown to have a strong synergism on filter paper with E5, an endocellulase, and CBHI, an exocellulase (Irwin et al., 1993). The ability of each enzyme to give synergism in filter paper hydrolysis with E5 and CBHI was measured to identify the influence of the cellulose binding domain on the synergistic reactions. The results of these assays are given in Table 2 and Figure 5A, and they show that the enzymatic activities of the cloned *E*E3 on CMC, swollen cellulose, and filter paper are nearly identical to those of *T*E3. Assayed on CMC, *S*E3 had 10 times the activity of *T*E3 and *E*E3 and twice as much activity as *T*E3 and *E*E3 on filter paper. However, a CMC overlay of a native gel on *S*E3 (Figure 3C) clearly showed that *S*E3 contained a contaminating CMCase from *S. lividans*. In the synergistic reactions, *E*E3, *S*E3, and *T*E3 gave similar results, and the contaminating enzyme did not appear to affect the activity. The binding of the three types of E3 to Avicel was very similar (Figure 5B). However, E3cd bound much more weakly as was found for E5cd and E2cd (Irwin et al., 1993). E3cd retained 77% of the activity of E3 on CMC and swollen cellulose but only 33% of its activity on filter paper. These results agree well with the studies of CBHII and CBHIIcd activity (Tomme et al., 1988). When E3cd was substituted for intact E3 in synergistic mixtures, approximately 30−50% of the activity was retained in the combinations as shown in Table 2 and Figure 5A. These results suggest that unlike endocellulase E5, the binding domain of E3 is required for synergism.

*Thermostability and pH Optimum of E3.* Purified *T*E3, *E*E3, E3cd, and CBHII at a concentration of 5 $\mu$mol/L in storage buffer (0.05 M NaOAc, pH 5.5) were incubated for 16 h at temperatures from 10 to 70 °C, followed by activity assays on swollen cellulose for each untreated and treated sample (Figure 6A). The pH−activity profile of E3 is shown in Figure 6B.

*Glycosylation of E3 and E3cd.* In a previous study, *T. fusca* E3 and E2 were shown to be glycosylated. E2 was estimated to contain about 2% sugar, and E3 appeared to have about 5% sugar (Wilson, 1988). To investigate differences in glycosylation, the purified E3s were tested for glycosylation (Figure 4C), and *E*E3 was not glycosylated, *S*E3 was partially glycosylated, and *T*E3 was glycosylated. No sugar was detected in *T*E3cd.

*Comparison of the Amino Acid Sequence of E3 with Those of Other Cellulases.* The amino acid sequence of E3 was compared with other protein sequences in the GeneBank and Swiss-Prot computer databases. A significant level of homology was present in the catalytic domain of one exocellulase and five endocellulases which were previously identified in cellulase family B (Gilkes et al., 1991b). The E3cd sequence was aligned to each of them to assess their relative similarities by the BESTFIT program. The level of identity and similarity for E3 was highest with *T. reseei* CBHII, 35% and 60% (Rouvinen et al., 1990), compared to 27% and 53% for *T. fusca* E2 (Lao et al., 1991). The four aspartate residues ($Asp^{175}$, $Asp^{221}$, $Asp^{263}$, and $Asp^{401}$) involved in the catalytic site of CBHII (Rouvinen et al., 1990) are conserved in all the enzymes in family B, suggesting that both the exo- and endocellulases share the same general catalytic mechanism for cleavage of the glycoside linkage. X-ray diffraction studies on E2cd also support the above hypothesis (Spezio et al., 1993).

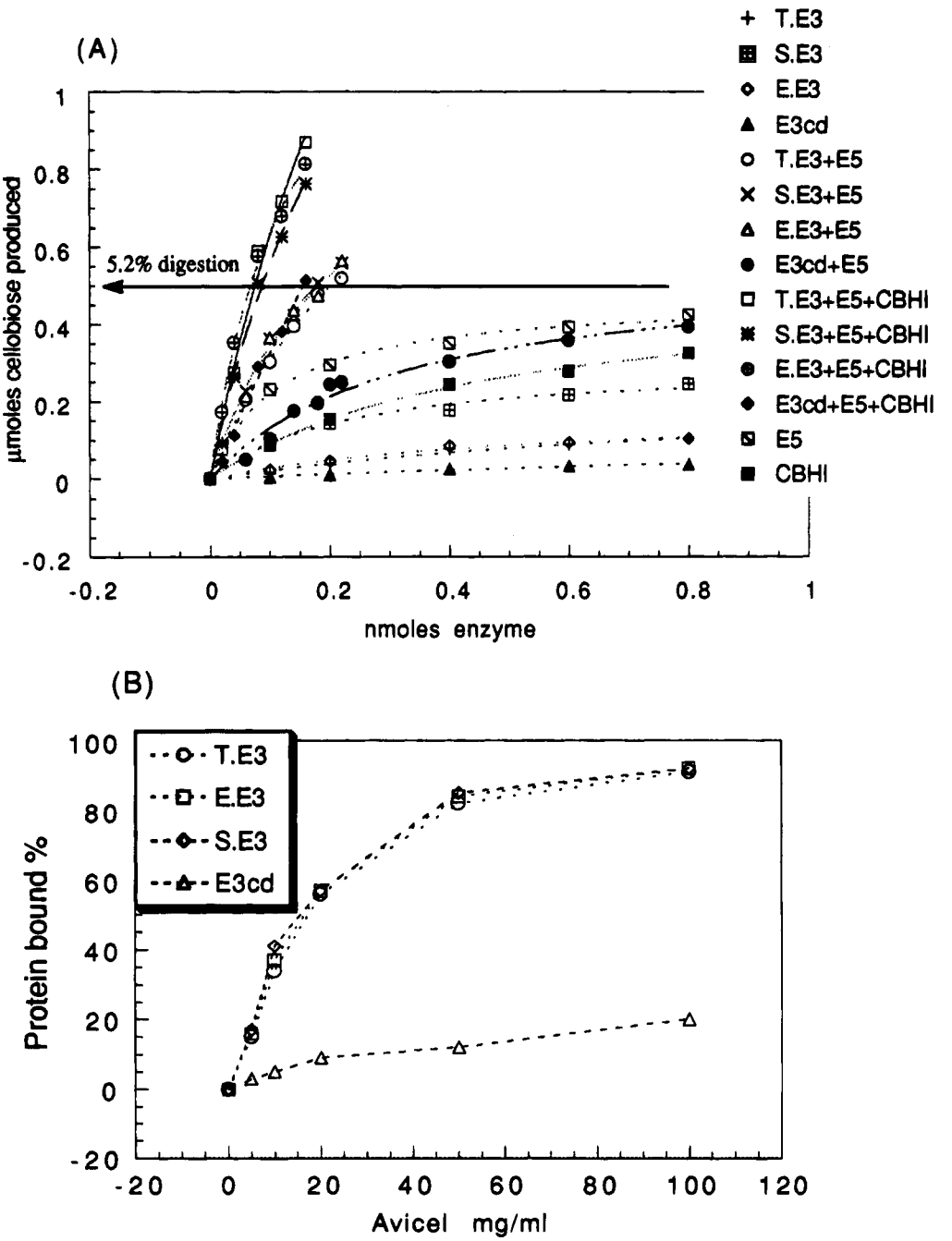


FIGURE 5: (A) Filter paper activity of E3 and E3cd. Assays are described under Materials and Methods. (B) Binding of E3 and E3cd to Avicel as described under Materials and Methods.

An unexpected difference is $Ile^{273}$ in E3 instead of the Pro found in all other family members, since this Pro is in the Leu/Ile-Glu-Pro region that was previouly proposed to be the Asn-Glu-Pro region present in 16 other cellulases and involved in catalysis (Gilkes et al., 1991a).

The four cysteines were reported to form two disulfide bonds to help stabilize two loop structures to form the tunnel (Rouvinen et al., 1990). These disulfide bridges were shown to be conserved in two endocellulases of the same family, E2cd (McGinnis & Wilson, 1993) and CenA of *C. fimi* (Gilkes et al., 1991a), and all seven family members possess the four cysteines at relatively similar positions.

A significant difference between the exo- and endocellulases is two loops corresponding to 172−189 and 394−429 in CBHII. Rouvinen et al. suggested that the basic difference between the exo- and the endocellulases is closure of the CBHII active site due to the two large loops which are deleted in the endocellulases (Rouvinen et al., 1990). Spezio confirmed the difference in the active site enclosure, and showed that the loop (corresponding to CBHII 172-189) in E2cd is not completely deleted. The five amino acid deletion in this loop shortens the helix and makes the loop pull back from the top of the cleft giving an open active site rather than the tunnel seen in CBHIIcd (Spezio et al., 1993). There are several additions to both loops in E3.

A typical proline-rich linker sequence was found to separate E3cd from the binding domain; it probably contains 30 residues from residues 106 to 135. Despite the sequence identity seen between linkers found in cellulases from the same organism (Gilkes et al., 1991b), the E3 linker is quite different from the other *T. fusca* linkers.

The E3 cellulose binding domain (CBD) was compared to other protein sequences using the FASTA search and showed high identity to several cellulase CBDs. It gave 58%

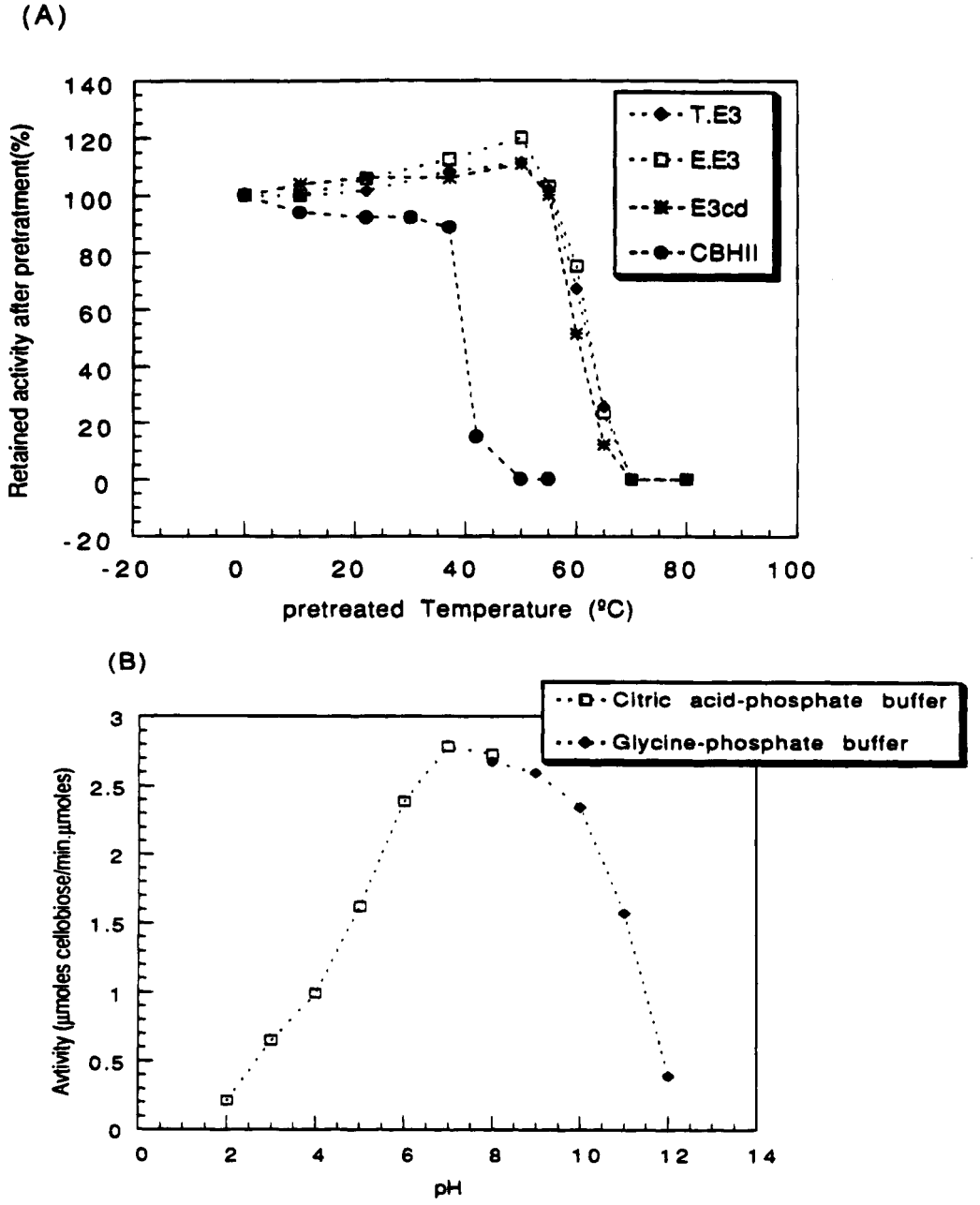


FIGURE 6: (A) Comparison of the thermostability of *T. fusca* E3 and *T. reseei* CBHII. (B) Activity of E3 on swollen cellulose at different pH values in citric acid–phosphate buffer and glycine-phosphate buffer.

identity to *C. fimi* CenA CBD (Wong et al., 1986) and 55% identity to the *C. fimi* Cex CBD (O'Niell et al., 1986). Recently it was reported that two Trp mutants (W14A and W68A) in the *C. fimi* CenA binding domain dramatically reduced their adsorption to crystalline cellulose by 50- and 30-fold, respectively (Din et al., 1994). Both residues are conserved in the E3 CBD.

## DISCUSSION

There has been a considerable amount of controversy concerning the production of exocellulases by bacteria. However, the studies reported here provide conclusive evidence not only that E3 is a thermostable exocellulase but also that it has nearly the same active site as *T. reesei* CBHII, a well-studied fungal exocellulase.

There have been a number of studies of the role of glycosylation in cellulase function (Goyal et al., 1991; Klyosov, 1990; Van Arsdell et al., 1987). In the case of E3, the fact that we cannot see any difference in the enzymatic activity, cellulose binding, thermostability, or proteolytic sensitivity between *T*E3, *S*E3, and *E*E3 demonstrates that bound sugars are not required for any of these functions since *E*E3 lacks bound sugars. It is also unlikely that bound sugars are needed for secretion since only two of six *T. fusca* cellulases contained bound sugar and all six are secreted.

The sum of the masses of the *T*E3 cyanogen bromide fragments is slightly larger than that of *T*E3. The difference between the *E*E3 mass and the *T*E3 mass, 1550 Da, suggests that, on average, nine hexose residues are present per molecule in *T*E3, and the 900 Da of extra mass in *S*E3 indicates that *S*E3 possesses five hexose residues. It is surprising that E3 was less degraded during its production by *S. lividans* than the four other *T. fusca* cellulases we have expressed in *S. lividans*. One possible explanation is that the E3 linker region where proteolysis appears to occur is

different from the linker regions of the other cellulases. However, it was possible to produce an E3 catalytic domain by treatment of E3 with papain. The N-terminal sequence of E3cd showed that all of the binding domain and almost all of the linker peptide were cleaved off by papain.

The E3 catalytic domain is very similar to CBHIIcd (76% similarity), so it is likely that the three-dimensional structure of E3cd is very close to that of CBHIIcd (Rouvinen et al., 1990). This also explains why the enzymatic properties of E3 are similar to those of CBHII (Irwin et al., 1993). E3 contains six cysteine residues, two in the binding domain and four in the catalytic domain. Because of the conservation of the cysteine residues in family B (Gilkes et al., 1991a), it is very likely that $Cys^{227}$ is joined to $Cys^{292}$ in one disulfide bond and $Cys^{465}$ is joined to $Cys^{515}$ in another while $Cys^{3}$ and $Cys^{101}$ form a third bond within the binding domain.

The distance from the transcriptional start point to the initiation codon in the E3 gene was within the range of 9−345 nucleotides found for 48 *Streptomyces* genes (Strohl, 1992). However, it was longer than the 104−116 bp of the corresponding regions in the E5 gene (Lin & Wilson, 1988); this long 5′ untranslated sequence contains putative regulatory sequences and potential secondary structure. Comparison of the level of expression of E3 in *S. lividans* transformed by pSZ7, pSZ7A, pSZ8, and pSZ11/12 which contain the E3 gene with the E3 promoter shows that pSZ7A has twice the activity of pSZ7 and 10-fold more than pSZ8 and produces as much E3 as pSZ11/12. This suggests that the *Tsr* gene promoter which is the closest promoter in pIJ702 to the E3 gene in pSZ7A can be as efficiently utilized to express E3 in *S. lividans* as the E3 promoter. The fact that the orientation of the *Tsr* promoter had no significant effect on the expression of E3 transformants containing the E3 promoter suggests that *S. lividans* recognizes the E3 promoter as found previously for E4 (Jung et al., 1993). pSZ7A which lacks the pUC18 fragment appeared to be more stable than strains that contained the shuttle plasmid.

## ACKNOWLEDGMENT

We gratefully acknowledge the technical assistance of Diana Irwin.

## REFERENCES

Bethesda Research Laboratory (1986) *Focus 8*, 9.
Burmeister, M., & Lehrach, H. (1989) *Trends Genet. 5*, 41.
Calza, R. E., Irwin, D. C., & Wilson, D. B. (1985) *Biochemistry 24*, 7797−7804.
Chung, C. T., Niemela, S. L., & Miller, R. H. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2172−2175.
Collmer, A., & Wilson, D. B. (1983) *Bio/Technology 1*, 594−601.
Dale, B. E. (1987) *Trends Biotechnol. 5*, 387−391.
Din, N., Forsythe, I. J., Burtnick, L. D., Gilkes, N. R., Miller, R. C. J., Warren, R. A. J., & Kilburn, D. G. (1994) *Mol. Microbiol. 11*, 747−755.
Dower, W. J., Miller, J. F., & Ragsdale, C. W. (1988) *Nucleic Acids Res. 16*, 6127−6145.
Ghangas, G. S., & Wilson, D. B. (1987) *Appl. Environ. Microbiol. 53*, 1470−1475.
Ghangas, G. S., & Wilson, D. B. (1988) *Appl. Environ. Microbiol. 54*, 2521−2526.
Ghangas, G. S., Hu, Y., & Wilson, D. B. (1989) *J. Bacteriol. 171*, 2963−2969.
Gilkes, N. R., Warren, R. A. J., Miller, R. C., & Jr & Kilburn, D. G. (1988) *J. Biol. Chem. 263*, 10401−10407.
Gilkes, N. R., Claeyssens, M., Aebersold, R., Henrissat, B., Meinke, A., Morrison, D. H., Kilburn, D. G., Warren, R. A. J., & Miller, R. C., Jr. (1991a) *Eur. J. Biochem. 202*, 367−377.
Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr., & Warren, R. A. J. (1991b) *Microbiol. Rev. 55*, 303−315.
Goyal, A., Ghosh, B., & Eveleigh, D. (1991) in *Bioresource Technology*, pp 37−50, Elsevier Science Publishers Ltd., New York.
Hanahan, D. (1983) *J. Mol. Biol. 166*, 557−580.
Hopwood, D. A., et al. (1985) in *Genetic Manipulation of Streptomyces—A Laboratory Manual*, The John Innes Foundation, Norwich, England.
Hu, Y. J., & Wilson, D. B. (1988) *Gene 71*, 331−337.
Hutter, R., & Eckhardt, T. (1988) in *Actinomycetes in biotechnology* (Goodfellow, M., Williams, S. T., & Mordarski, M., Eds.) pp 89−184, Academic Press, Inc., New York.
Irwin, D. C., Spezio, M., Walker, L., & Wilson, D. B. (1993) *Biotechnol. Bioengin. 42*, 1002−1013.
Jung, E. D., Lao, G., Irwin, D., Barr, B. K., Benjamin, A., & Wilson, D. B. (1993) *Appl. Environ. Microbiol. 59*, 3032−3043.
Kinston, R. E. (1989) in *Current Protocols in Molecular Biology* (Ausubel, F. M., et al., Eds.)John Wiley & Sons, New York.
Klyosov, A. A. (1990) *Biochemistry 29*, 10577−10585.
Laemmli, U. K. (1970) *Nature (London) 227*, 680−685.
Lao, G., Ghangas, G. S., Jung, E. D., & Wilson, D. B. (1991) *J. Bacteriol. 173*, 3397−3407.
Lin, E., & Wilson, D. B. (1988) *J. Bacteriol. 170*, 3838−3842.
Lynd, L. R., Cushman, J., Nichols, R., & Wyman, C. (1991) *Science 251*, 1318−1323.
Matsudaira, P. (1987) *J. Biol. Chem. 262*, 10035−10038.
McGinnis, K., & Wilson, D. B. (1993) *Biochemistry 32*, 8151−8156.
O'Neill, G. P., Goh, S. H., Warren, R. A. J., Kilburn, D. G., & Miller, R. C., Jr. (1986) *Gene 44*, 325−330.
Petricek, M., Stajner, K., & Tichy, P. (1989) *J. Gen. Microbiol. 135*, 3303−3309.
Rosenberg, M., & Court, D. (1979) *Annu. Rev. Genet. 13*, 319−353.
Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., & Jones, T. A. (1990) *Science 249*, 380−386.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanger, F., Nicklon, S., & Conlson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
Spezio, M., Wilson, D. B., & Karplus, A. (1993) *Biochemistry 32*, 9906−9916.
Strohl, W. R. (1992) *Nucleic Acids Res. 20*, 961−974.
Summers, W. C. (1970) *Anal. Biochem. 33*, 459−463.
Teather, R. M., & Wood, P. J. (1982) *Appl. Environ. Microbiol. 43*, 777−780.
Tomme, P., et al. (1988) *Eur. J. Biochem. 170*, 575−581.
Towbin, H., Staehelin, T., & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350−4354.
Van Arsdell, J. N., Kwok, S., Schweickart, V., Ladner, M. B., Gelfand, D. H., & Innis, M. A. (1987) *Bio/Technology 5*, 60−64.
Van Tilbeurgh, H., Bhikhabbhai, R., Pettersson, L. G., & Claeyssens, M. (1984) *FEBS Lett. 169*, 215−218.
Van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R., & Pettersson, G. (1986) *FEBS Lett. 204*, 223−227.
Villa, S., DeFazio, G., & Canosi, U. (1989) *Anal. Biochem. 177*, 161−164.
Wilson, D. B. (1988) *Methods Enzymol. 160*, 314−323.
Wilson, D. B. (1992) *Crit. Rev. Biotechnol. 12*, 45−63.
Wong, W. K. R., Gerhark, B., Guo, Z. M., Kilburn, D. G., Warren, R. A. J., & Miller, R. C., Jr. (1986) *Gene 44*, 315−324.

BI941789R