Cloning, sequencing and homologies of the *cbh-1* (exoglucanase) gene of *Humicola grisea* var. *thermoidea*

MARISTELLA DE O. AZEVEDO,¹† MARIA S. S. FELIPE,²† SPARTACO ASTOLFI-FILHO²†
and ALAN RADFORD¹*

¹Department of Genetics, The University of Leeds, Leeds LS2 9JT, UK

²Department of Biochemistry, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK

(Received 5 March 1990; revised 30 July 1990; accepted 20 August 1990)

Studies on the enzymes of the cellulase complex of the thermophilic fungus *Humicola grisea* var. *thermoidea* are described. A genomic library was constructed in the phage vector EMBL 4, and from this library two clones were isolated using as a probe the cloned *cbh-1* (exoglucanase, EC 3.2.1.91) gene of *Phanerochaete chrysosporium*, a cellulolytic basidiomycete fungus. These clones were analysed by restriction mapping and Southern blotting, and one of them (λ 3) was sub-cloned into the M13 phage vectors mp18 and mp19. The gene sequence was determined by the dideoxy chain-termination method. Sequence comparison with the equivalent genes from *P. chrysosporium* and *Trichoderma reesei* was made: in terms of primary sequence there is about 60% homology between the three species. Secondary structure prediction of the *H. grisea* sequence was also computed.

Introduction

Humicola grisea var. *thermoidea* is a thermophilic imperfect filamentous fungus. The isolate used in this work is an efficient cellulolytic organism, producing high levels of the enzymes of the cellulase complex, namely endoglucanases (EC 3.2.1.4), exoglucanases (EC 3.2.1.91) and β -glucosidase (EC 3.2.1.21) (Araujo *et al.*, 1983; R. Monti, personal communication). The enzymes of the cellulase complex have commercial potential as well as considerable scientific interest. Cellulolytic enzymes have been studied in other species of fungi, such as *Trichoderma reesei* (Shoemaker *et al.*, 1983; Teeri *et al.*, 1983), *Phanerochaete chrysosporium* (Sims *et al.*, 1988) and *Neurospora crassa* (Eberhardt, 1961; Eberhardt & Beck, 1973; Yazdi *et al.*, 1990a, b). Some of the cellulase complex genes of *T. reesei* and *P. chrysosporium* have been cloned and sequenced.

In general, the endoglucanases and exoglucanases are exported enzymes because of the insolubility of their substrate, cellulose, but the β -glucosidases, which further

degrade the cellobiose resulting from the action of the glucanases, are predominantly intracellular (Gong & Tsao, 1980). Endoglucanases and exoglucanases have a characteristic domain structure, with the greater part of the sequence specifying a catalytic domain, separated from a cellulose-binding domain by a 'hinge' rich in serine or proline and threonine residues (Knowles *et al.*, 1987).

Here we describe the analysis, by gene cloning and sequencing, of an enzyme of the cellulase complex of *H. grisea* var. *thermoidea*.

Methods

Fungal strain. The strain of *H. grisea* var. *thermoidea* used was a wild isolate obtained in Brazil (Araujo *et al.*, 1983).

It was routinely maintained on slants of 4% (w/v) oat flour, 1.5% (w/v) agar, on which good conidiation was achieved after 5 d at 42 °C. For the production of mycelium, a rich liquid medium was used, based on the minimal medium of Pontecorvo *et al.* (1953), supplemented with 0.2% peptone, 0.15% hydrolysed casein, yeast extract, yeast nucleic acid and vitamin solution, adjusted to pH 6.8 (Azevedo & Costa, 1973).

Phage vectors and bacterial strains. The phage vector in which the library was constructed was EMBL4 (Frischauf *et al.*, 1983). Sub-cloning for DNA sequencing was into the M13 phage vectors mp18 and mp19 (Messing, 1983; Yanisch-Perron *et al.*, 1985). *Escherichia coli* host strains used were TG1, L2392 (Maniatis *et al.*, 1982) and P2392, a derivative of L2392 (Perbal, 1989).

† Present address: Department of Cellular Biology, The University of Brasilia, Brasilia, DF, Brazil.

The nucleotide sequence data reported in this paper have been submitted to EMBL and have been assigned the accession number X17258.

DNA extraction. Mycelium was grown overnight at 42 °C in the rich liquid medium from a large inoculum of conidia (about 10^5 conidia ml^{-1}) from a 7–14-d-old culture. The mycelium was harvested by filtration, and stored at –80 °C. A 2 g aliquot of mycelium was ground under liquid nitrogen, resuspended in 15 ml TTE buffer (10 mM-Tris pH 8.5, 10 mM-EDTA, 4 mM-spermidine, 10 mM-2-mercaptoethanol, 0.5 M-sucrose, 36 mM-KCl, 0.25% Triton X-100), and centrifuged at 2500 g for 10 min. The pellet was carefully resuspended in lysis buffer (40 mM-Tris pH 8.0, 10 mM-EDTA, 0.2 M-NaCl, 1.5% w/v, SDS), and extracted with double-distilled phenol liquified with buffer (Maniatis *et al.*, 1982). The aqueous phase was extracted a second time with phenol, and then with chloroform. Then 0.1 vol. 3 M-sodium acetate (pH 5.5) was added, and the mixture was ethanol-precipitated at –20 °C. The precipitated DNA was spooled out with a glass rod, washed with 70% (v/v) ethanol, dried at room temperature, and redissolved in TE buffer (10 mM-Tris, 10 mM-EDTA; pH 8.0) to a final concentration of about 0.5 $\mu\text{g}/\mu\text{l}^{-1}$.

Construction and screening of the genomic library. The *H. grisea* genomic library was constructed by completely digesting the λ phage vector EMBL4 (Promega Biotech) with *Bam*HI and *Sal*I, and making a partial digest of *H. grisea* genomic DNA with *Sau*3A to an average fragment length of 15 kb. Fragments in the size range 9–18 kb were ligated into the cut vector, the ligation mixture packaged with a λ packaging kit (Gigapack) according to the manufacturer's instructions, and the packaged mixture plated onto *E. coli* L2392 (Maniatis *et al.*, 1982). The gene library was amplified in *E. coli* P2392.

The library was screened by plaque hybridization, using *E. coli* P2392 as host. The probe employed was a 1.7 kb *Pst*I genomic fragment containing a clone of the *Phanerochaete chrysosporium cbh-1* gene (Sims *et al.*, 1988). The probe was labelled with [α - ^{32}P]dATP, using a random primer kit (Boehringer Mannheim) according to the manufacturer's instructions.

After plating an aliquot of the library on P2392 according to Maniatis *et al.* (1982), and incubating overnight at 37 °C, the plates were blotted onto nitrocellulose filters (Schleicher and Schuell), denatured for 2 min, and then neutralized. The filters were then washed, baked, incubated with pre-hybridization solution, hybridized with the probe and analysed by autoradiography. Plaque areas giving a good positive signal were identified, replated at lower plaque density, and individual positive plaques identified and isolated (Maniatis *et al.*, 1982).

Positive clones were further investigated by restriction and Southern blotting. DNA from these was prepared, and cut with *Eco*RI, *Bam*HI and *Sal*I, singly and in double digests, and run on agarose gels, with standard size markers (*Hind*III-cut λ and *Hae*III-cut ϕX174). After denaturing the DNA, the gels were blotted onto nitrocellulose membranes (Schleicher and Schuell), transferring overnight in $10\times$ SSC ($1\times$ SSC is 0.15 M-NaCl, 0.015 M-sodium citrate, pH 7.0). The membranes were baked, incubated in pre-hybridization solution, incubated with probe, washed, and analysed by autoradiography.

Sub-cloning and sequencing. For sub-cloning into M13 vectors mp18 and mp19, vector DNA was digested with *Eco*RI/*Sal*I. Either total or agarose-gel-purified clone DNA was then digested with *Eco*RI/*Sal*I and ligated into the cut vector DNA. Competent cells of *E. coli* TG1 were prepared according to Chung *et al.* (1989) or Mandel & Higa (1970), and transformed with the ligation mixture according to the method of Chung *et al.* (1989). After transformation and plating on medium containing IPTG and X-Gal, white plaques were isolated. From these, replicative-form DNA was prepared, digested, run on an agarose gel, and probed with the *P. chrysosporium cbh-1* clone to identify those sub-clones containing inserts of interest. Sequencing was done by using the dideoxy chain-termination method (Sanger *et al.*, 1977) and T7 DNA polymerase, in kit form (Pharmacia).

The software of Staden (1982) was used to compile and translate the gene sequence. Alignment of the *H. grisea* sequence with those of the equivalent *T. reesei* and *P. chrysosporium* genes was carried out with MANALIGN, one of a suite of sequence analysis programs in the OWL database and associated suite of analytical software (Akriag *et al.*, 1988). Secondary structure was computed with PREDICT, another program written for the OWL database, which integrates the methods of both Chou & Fasman (1978) and Garnier *et al.* (1976).

Results and Discussion

A *Sau*3A partial digest of *H. grisea* genomic DNA of size range 9–18 kb was ligated into *Bam*HI/*Sal*I-cut EMBL4 vector DNA. This ligation mixture was mixed with cells of strain P2392, and plated. A total of 3×10^6 plaques were obtained from 0.5 ml of the packaged mixture, harvested and stored. For an organism with a genome of about 10^7 bp and inserts of 10–15 kb, coverage of the genome in the library at a probability of 0.99 requires 15000 clones.

To screen the *H. grisea* genomic library with the *P. chrysosporium cbh-1* probe, approximately 6×10^4 plaques on a total of six 9 cm diameter plates were blotted in duplicate onto nitrocellulose filters, and probed with [^{32}P]dATP-labelled isolated insert. Thirty positive plaques were detected. These were isolated, replated at low plaque density, and re-screened. Two of them, identified as $\lambda 3$ and $\lambda 9$, gave clear and uniform positive hybridization in this re-screening, and were selected for further study.

Restriction mapping of clones $\lambda 3$ and $\lambda 9$ showed them to be different in their restriction patterns with *Eco*RI,

Fig. 1. Restriction maps of the *H. grisea* clones $\lambda 3$ and $\lambda 9$. In each case the bold line indicates the region to which the *P. chrysosporium* probe, essentially the entire coding region of the *cbh-1* gene, hybridized. The restriction sites indicated are those spanning the fragments to which the probe bound, and hence the closest sites for *Eco*RI (E), *Sal*I (S) and *Bam*HI (B) on either side of the gene. Distances are given in kb.

Fig. 2. The sequencing strategy employed for *H. grisea cbh-1*. The figure indicates the total length sequenced, the positions of the intron and (bold lines) two exons, and the location of the *SalI* site in the second exon. The arrows below indicate the positions of sequencing primers, and the direction and extent of sequence determined from each. For the initial sequences out from the *SalI* site, M13 '-40' primer was used. (⇒ specific primers).

*Bam*HI and *Sal*I. The restriction gel was blotted, and probed with the isolated *P. chrysosporium cbh-1* clone labelled with [32 P]dATP. With one exception, this gave a single hybridizing band with each single restriction digest. In the case of clone λ 3 cut by *Sal*I, two bands hybridized, indicating a restriction site within the region of λ 3 to which the probe was hybridizing. As the *P. chrysosporium* probe contained little other than the coding region of the *cbh-1* gene, this indicated a probable *Sal*I site within an equivalent gene in λ 3. Fig. 1 illustrates restriction maps of clones λ 3 and λ 9, based on the sizes of the hybridizing restriction fragments.

It is clear from the data in Fig. 1 that clones λ 3 and λ 9 are different in their restriction patterns, but they show approximately equal levels of similarity to the probe, based on the efficiency of hybridization. As is shown below, λ 3 contains a gene that is clearly homologous to the *cbh-1* genes of *T. reesei* and *P. chrysosporium*. λ 9 has not yet been further characterized, but its high level of hybridization to the probe suggests that it may well contain another gene of the cellulase complex. In support of this conclusion, the homology between the *cbh-1* (exo-glucanase) and *eg-1* (endo-glucanase) genes of *T. reesei* is quite high in parts of their sequence, high enough to use one as a probe for the other.

Based on the above restriction and hybridization data, *Eco*RI/*Sal*I double digest fragments of λ 3 were sub-cloned into mp18 and mp19, the ligation mixture transformed into *E. coli* strain TG1, and sub-clones detected as white plaques on X-Gal plates. Inserts of interest were identified by probing with the *P. chrysosporium cbh-1* fragment.

As the restriction analysis had shown an internal *Sal*I site, λ 3 sub-clones which had a *Sal*I site adjacent to the primer site were chosen for sequencing. From this initial sequencing, reading was extended further with custom oligonucleotide primers until the coding region had been traversed. Complementary sub-clones were then used to read back on the second strand to, and through, the

starting *Sal*I site. The entire sequencing strategy is shown in Fig. 2.

Data from the first sequencing gels, reading out from the internal *Sal*I site, were translated in all six possible reading frames, and the homology with the *T. reesei* and *P. chrysosporium cbh-1* gene sequences at the amino-acid level was immediately apparent. This identified unequivocally the correct reading frame, orientation and approximate location of the sequenced region within the gene. Oligonucleotides (17-mers) corresponding to the furthest unambiguous sequence from each reading were synthesized, and used as primers to extend the reading on the same sub-clones. This process was repeated until the entire open reading frame and beyond had been sequenced. Complementary 17-mer primers were used for sequencing the second strand back to the internal *Sal*I site, using further appropriate sub-clones. Altogether 2.046 kb of sequence was determined (Fig. 3).

Analysis of the DNA sequence of λ 3, and computer translation of a large open reading frame (Azevedo & Radford, 1990), together with preliminary comparisons with the *cbh-1* genes of *T. reesei* and *P. chrysosporium* revealed a coding region of 1575 nucleotides, containing two exons interrupted by a single, small intron of 60 nucleotides. The predicted translation product of the open reading frame would have an M_r of 55654, but this does not take into account the effect of probable glycosylation of such an extracellular enzyme. The first exon starts with a typical N-terminal signal sequence of about 18 amino acid residues, with the basic residues arginine and lysine near the N-terminus followed by a short region rich in hydrophobic residues. The intron is of a length typical for filamentous ascomycete nuclear gene introns. It is defined by the 5' consensus GTA and the 3' consensus CAG, and contains, starting 21 nucleotides upstream from the 3' end, the internal consensus GCTAAC. This intron does not correspond in position to those found in the equivalent gene of the other two species. If it were to be translated, this intron would

gccgtgaccttgcgcgctttgggtggcggtggcgagtcgtggacgggtgcttgctgggtcgccggccttccc 70
 ggcgatccgctgatgagagggccaccaacggcggggatgatgctccatggggaacttccccatggagaag 140
 agagagaaacttgcgagccgtgatctggggaaagatgctccgtgtctcgtctatataaactcgagtctcc 210
 ccgagccctcaacaccaccagctctgatctcaccatccccatcgacaatcacgcaaacacagcagttgtc 280
 M R T A K F A T L A
 gggccatttccttcagacacatcagtcaccctccttcaaaaATGCGTACCGCCAAGTTCGCCACCCTCGCCG 350
 A L V A S A A A Q Q R C S L T T E R H P S L S W
 CCCTTGTGGCCTCGGCCGCCGCCAGCAGGCGTGCAGTCTCACCACCGAGAGGCACCCTTCCCTCTCTTG 420
 K K C T A G G Q C Q T V Q A S I T L D S N W R
 GAACAAGTGCACCGCCGGCGGCCAGTGCCAGACCGTCCAGGCTTCCATCACTCTCGACTCCAACCTGGCGC 490
 W T H Q V S G S T N C Y T G N K W D T S I C T
 TGGACTCACCAGGTGTCTGGCTCCACCAACTGCTACACGGGCAACAAGTGGATACTAGCATCTGCACTGA 560
 D A K S C A H N C C V D G A D Y T S T Y G I T T
 TGCCAAGTTCGTGCGCTCAGAACTGCTGCGGTGATGGTGCCGACTACACCAGCACCTATGGCATCACCAC 630
 N G D S L S L K F V T K G Q H S T N V G S R T
 CAACGGTGATTCCCTGAGCCTCAAGTTCGTACCAAGGGCCAGCACTCGACCAACGTTCGGCTCGCGTACC 700
 Y L M D G E D K Y Q
 TACCTGATGGACGGCGAGGACAAGTATCAGAgtaagttctatcttcagccttctcgcgccttgaatcctg 770
 T F E L V G N E F T F D V D V S N
 gctaacgtttacacttcacagCCTTCGAGCTCCTCGGCAACGAGTTCACCTTCGATGTGATGTCTCCAA 840
 I G C Y L N G A L Y F V S M D A D G G L S R Y
 CATCGGCTGCGGTCTCAACGGCGCCCTGTACTTCGTCTCCATGGACGCCGATGGTGGTCTCAGCCGCTAT 910
 P G N K A G A K Y G T G Y C D A Q C P R D I K
 CCTGGCAACAAGGCTGGTGCCAAGTACGGTACCGGCTACTGCGATGCTCAGTGCCCCCGTGACATCAAGT 980
 F I N G E A N I E G W T G S T N D P N A G A G R
 TCATCAACGGCGAGGCCAACATTGAGGGCTGGACCGGCTCCACCAACGACCCCAACGCCGGCGGGGCCG 1050
 Y G T C C S E M D I W E A Q Q H A T A F I P H
 CTATGGTACCTGCTGCTCTGAGATGGATATCTGGGAAGCCAACAACATGGCTACTGCCTTCACTCCTCAC 1120
 P C T I I G Q S R C E G D S C G G T Y S N E R
 CCTTGCACCATCATTGGCCAGAGCCGCTGCGAGGGCGACTCGTGCGGTGGCACCTACAGCAACGAGCGCT 1190
 Y A G V C D P D G C D F N S Y R Q G N K T F Y G
 ACGCCGGCGTCTGCGACCCCGATGGCTGCGACTTCAACTCGTACCGCCAGGGCAACAAGACCTTCTACGG 1260
 K G M T V H T T K K I T V V T Q F L K D A N G
 CAAGGGCATGACCGTGCACACCACCAAGAAGATCACTGTCTGTCACCCAGTTCCTCAAGGATGCCAACGGC 1330
 D L G E I K R F Y V Q D G K I I P N S E S T I
 GATCTCGGCGAGATCAAGCGCTTCTACGTCCAGGATGGCAAGATCATCCCCAACTCCGAGTCCACCATCC 1400
 P G V E G N S I T Q D W C D R Q K V A F G D I D
 CCGGCGTTCGAGGGCAATTCCATCACCCAGGACTGGTGCGACCGCCAGAAGGTTGCCTTTGGCGACATTGA 1470
 D F N R K G G M K Q M G K A L A G P M V L V M
 CGACTTCAACCGCAAGGGCGGCATGAAGCAGATGGGCAAGGCCCTCGCCGGCCCCATGGTCTGGTCATG 1540
 S I W D D H A S N M L W L D S T F P V D A A G
 TCCATCTGGGATGACCACGCCTCCAACATGCTCTGGCTCGACTCGACCTTCCCTGTGATGCCGCTGGCA 1610

K P G A E R G A C P T T S G V P A E V E A E A P
 AGCCCGGCGCGAGCGCGGTGCCCTGCCCCGACCACCTCGGGTGTCCCTGCTGAGGTTGAGGCCGAGGCCCC 1680
 N S N V V F S N I R F G P I G S T V A G L P G
 CAACAGCAACGTCGTCTTCTCCAACATCCGCTTCGGCCCCATCGGCTCGACCGTTGCTGGTCTCCCCGGC 1750
 A G N G G N N G G N P P P P T T T T S S A P A
 GCGGGCAACGGCGGCAACAACGGCGGCAACCCCCCGCCCCCACCACCACCTCCTCGGCTCCGGCCA 1820
 T T T T A S A G P K A G R W Q Q C G G I G F T G
 CCACCACCACCGCCAGCGCTGGCCCCAAGGCTGGCCGCTGGCAGCAGTGC GGCGGCATCGGCTTCACTGG 1890
 P T Q C E E P Y I C T K L N D W Y S Q C L
 CCCGACCCAGTGCAGGAGCCCTACATTTCACCAAGCTCAACGACTGGTACTCTCAGTGCCTGtaaatt 1960
 ctgagtcgctgactcgacgatcacggccggtttttgcatgaaaggaaacaaacgaccgcgataaaaaatgg 2030
 agggtaatgagatgtc

Fig. 3. The determined DNA sequence of the *H. grisea cbh-1* gene, and the translation of the exons. Also indicated, by underlining, are a possible TATA box (193–199), the internal *SalI* site from which sequencing commenced (771–776), and the consensus 5', 3' and internal sequence features of the intron (732–734; 789–791).

Table 1. Codon usage in the *cbh-1* gene of *H. grisea*

Phe	UUU	1	Ser	UCU	4	Tyr	UAU	4	Cys	UGU	0
	UUC	17		UCC	15		UAC	13		UGC	24
Leu	UUA	0		UCA	0	–	UAA	1	–	UGA	0
	UUG	0		UCG	10	–	UAG	0	Trp	UGG	11
	CUU	1	Pro	CCU	6	His	CAU	0		CGU	3
Leu	CUC	16		CCC	15		CAC	6	Arg	CGC	12
	CUA	0		CCA	0	Gln	CAA	0		CGA	0
	CUG	5		CCG	4		CAG	20		CGG	0
Ile	AUU	4	Thr	ACU	8	Asn	AAU	1	Ser	AGU	1
	AUC	18		ACC	44		AAC	32		AGC	8
	AUA	0		ACA	0	Lys	AAA	0	Arg	AGA	0
Met	AUG	11		ACG	1		AAG	24		AGG	1
	GUU	3	Ala	GCU	11	Asp	GAU	15		GGU	13
Val	GUC	18		GCC	32		GAC	18	Gly	GGC	50
	GUA	0		GCA	0	Glu	GAA	1		GGA	0
	GUG	3		GCG	3		GAG	19		GGG	0

give no amino acid sequence equivalent to any part of the sequence in the other species, and in fact would result in chain termination at an in-phase UAA codon.

Codon usage in the *H. grisea cbh-1* gene is shown in Table 1. It shows a marked inequality in frequency of bases in the third position of the codon, with a strong preference for uridine. For those amino acids for which it is necessary to have a purine base in the third position of the codon, there is a very strong preference for guanine. Although codon usage in nuclear genes of filamentous fungi in general is far from random (Kinghorn, 1988), the bias in this gene (Table 1) is probably more extreme than in any other to date, with third-position bases being 337 C, 111 G, 75 U but only 2 A. This bias against adenine in the third position is exceptional.

By comparison with other cellulases (Knowles *et al.*, 1987), the sequence of the *H. grisea cbh-1* product can be divided into domains, with a signal sequence from residues 1 to 18, an exoglucanase catalytic domain (19–456), a 'hinge' (457–487), and a cellulose-binding domain (488–525) (Fig. 4). The *H. grisea cbh-1* sequence was analysed using standard methods to determine hydrophathy and to make a secondary structure prediction (Fig. 4). The polypeptide has a hydrophobic N-terminal signal sequence as expected, but the remainder of the sequence is predominantly hydrophilic. Predicted α -helix forms the entire signal sequence, plus several short stretches in the catalytic domain, but is entirely absent from the hinge and cellulose-binding domain. Interspersed short blocks of β -sheet and turn coil form the major part of the

Fig. 4. Structural features of the *cbh-1* gene product. (a) The domain structure, indicating the export signal sequence (S), the catalytic domain, the hinge sequence (H), and the cellulose-binding domain (B). (b) Hydropathy plot, prepared by the methods of Kyte & Doolittle (1982). Above the horizontal line is hydrophobic, and below is hydrophilic. (c) Secondary structure plot indicating the relation to the catalytic, hinge and cellulose-binding domains of the enzyme in (a). □, predicted α -helix; ■, β -sheet; —, turn coil.

Table 2. Overall and domain sequence homology between the *cbh-1* gene of *H. grisea* and its equivalents in *P. chrysosporium* and *T. reesei*, plus the related *eg-1* endoglucanase of *T. reesei*

The signal sequence has been omitted from this table.

Gene	Domain:			Total
	Catalytic	Hinge	Binding	
<i>T. reesei cbh-1</i>	$\frac{250}{413}$ 61%	$\frac{17}{30}$ 57%	$\frac{20}{38}$ 53%	$\frac{287}{481}$ 60%
<i>P. chrysosporium cbh-1</i>	$\frac{251}{423}$ 59%	$\frac{1}{29}$ 3%	$\frac{21}{37}$ 57%	$\frac{273}{489}$ 56%
<i>T. reesei eg-1</i>	$\frac{251}{360}$ 44%	$\frac{6}{28}$ 21%	$\frac{17}{38}$ 45%	$\frac{273}{426}$ 43%

catalytic domain, and are the only structures predicted in the hinge and cellulose-binding domain. This same general arrangement of domains is also found in *cbh-1* in *T. reesei* and *P. chrysosporium* and in the *T. reesei eg-1* gene product. In the products of other cellulase complex genes in *T. reesei*, the cellulose-binding domain is N-terminal, and the catalytic domain C-terminal (Knowles *et al.*, 1987). Both patterns are found in bacterial cellulases. The catalytic domain-hinge-substrate-binding domain structure is found in certain other polysaccharide-digesting enzymes, e.g. amylases, and the hinge may similarly be rich in residues such as threonine, e.g. glucoamylase of *Aspergillus niger* (near the C-terminus) or *Saccharomyces cerevisiae* (near the N-terminus) (Svensson *et al.*, 1986; Yamashita *et al.*, 1985). In these examples also, the hinge is threonine-rich, but has less sequence homology to the *cbh-1* genes than they do between themselves.

The optimal alignment of the amino acid sequence of *H. grisea cbh-1* with its equivalents from *T. reesei* and *P. chrysosporium* was derived by using the program MANALIGN from the OWL Database software, running on a VAX/VMS computer, and is shown in Fig. 5. This figure also indicates the major domains within the sequence: N-terminal signal, hinge, and substrate-binding domain, with the long catalytic domain between the signal and the hinge. A breakdown of the homology for catalytic, hinge and cellulose-binding domains is shown in Table 2. In the hinge region, the *H. grisea* sequence resembles that of *T. reesei cbh-1* in being rich in proline and threonine residues, although these both differ markedly from the *P. chrysosporium cbh-1* hinge, which is rich in serine residues. Except for the hinge region, however, the three sequences are more or less equally related, with about 60% homology. This perhaps suggests that the proline/threonine-rich hinge of *H. grisea* and *T. reesei* is typical of

Fig. 5. Optimal alignment of the *H. grisea cbh-1* amino-acid sequence with those of the equivalent *cbh-1* sequences of *P. chrysosporium* and *T. reesei*, and also of the related *eg-1* (endoglucanase) sequence of *T. reesei*. The signal, hinge and substrate-binding domains are indicated, and the long catalytic domain is between the signal and the hinge. Asterisks and vertical lines denote identical residue in adjacent sequence, and identical residue in non-adjacent sequence, respectively.

[illegible]

this gene in ascomycetes, whereas the serine-rich hinge of *P. chrysosporium* is a basidiomycete characteristic.

This work was supported in part by CNPq, Brazil.

References

- AKRIGG, D., BLEASBY, A. J., DIX, N. I. M., FINDLAY, J. B. C., NORTH, A. C. T., PARRY-SMITH, D., WOOTTON, J. C., BLUNDELL, T. L., GADNER, S. P., HAYES, F., ISLAM, S., STERNBERG, M. J. E., THORNTON, J. M. & TICKLE, I. J. (1988). A protein sequence/structure database. *Nature, London* **335**, 745–746.
- ARAUJO, E. F., BARROS, E. G., CALDAS, R. A. & SILVA, D. O. (1983). Beta-glucosidase activity of a thermophilic cellulolytic fungus, *Hemicola* sp. *Biotechnology Letters* **5**, 781–784.
- AZEVEDO, J. L. & COSTA, S. O. P. (1973). *Exercícios Práticos de Genética*. São Paulo: Edições Universidade de São Paulo.
- AZEVEDO, M. DE O. & RADFORD, A. (1990). Sequence of the *cbh-1* gene of *Hemicola grisea* var. *thermoidea*. *Nucleic Acids Research* **18**, 668.
- CHOU, P. Y. & FASMAN, D. G. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology* **47**, 45–148.
- CHUNG, C. T., NIEMALA, S. L. & MILLER, R. H. (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 2172–2175.
- EBERHARDT, B. M. (1961). Exogenous enzymes of *Neurospora* conidia and mycelia. *Journal of Cellular and Comparative Physiology* **58**, 11–16.
- EBERHARDT, B. M. & BECK, R. S. (1973). Induction of β -glucosidases in *Neurospora crassa*. *Journal of Bacteriology* **116**, 295–303.
- FRISCHAUF, A. M., LEHRACH, H., POUSTKA, A. & MURRAY, N. (1983). Lambda replacement vectors carrying polylinker sequences. *Journal of Molecular Biology* **170**, 827–842.
- GARNIER, J., OSGUTHORPE, D. J. & ROBSON, B. (1976). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology* **120**, 97–120.
- GONG, C. S. & TSAO, G. T. (1980). Cellulase biosynthesis and regulation. *Annual Reports on Fermentation Processes* **3**, 111–140.
- KNOWLES, J., LEHTOVAARA, P. & TEERI, T. (1987). Cellulase families and their genes. *Trends in Biotechnology* **5**, 255–261.
- KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105–132.
- MANDEL, M. & HIGA, A. (1970). Calcium-dependent bacteriophage infection. *Journal of Molecular Biology* **53**, 159–162.
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning – a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- MESSING, J. (1983). New M13 vectors for cloning. *Methods in Enzymology* **101c**, 20–79.
- PERBAL, B. (1989). *A Practical Guide to Molecular Cloning*, 2nd edn. New York: Wiley.
- PONTECORVO, G., ROPER, J. A. & FORBES, E. (1953). Genetic recombination without reproduction in *Aspergillus niger*. *Journal of General Microbiology* **8**, 198–210.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.
- SHOEMAKER, S., SCHWEICKART, V., LADNER, M., GELFAND, D., KWOK, S., MYAMBO, K. & INNIS, M. (1983). Molecular cloning of exo-cellobiohydrolase 1 derived from *Trichoderma reesei* strain L27. *Bio/technology* **1**, 691–696.
- SIMS, P., JAMES, C. & BRODA, P. (1988). The identification, molecular cloning and characterization of a gene from *Phanerochaete chrysosporium* that shows strong homology to the exo-cellobiohydrolase I gene of *Trichoderma reesei*. *Gene* **74**, 411–422.
- STADEN, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Research* **10**, 2951–2961.
- SVENSSON, B., LARSEN, K. & GUNNARSSON, A. (1986). Characterisation of a glucoamylase G2 from *Aspergillus niger*. *European Journal of Biochemistry* **154**, 497–502.
- TEERI, T., SALOVUORI, I. & KNOWLES, J. (1983). The molecular cloning of the major cellulase gene from *Trichoderma reesei*. *Bio/technology* **1**, 696–699.
- YAMASHITA, I., SUZUKI, K. & FUKUI, S. (1985). Nucleotide sequence of the extracellular glucoamylase gene *STA1* in the yeast *Saccharomyces cerevisiae*. *Journal of Bacteriology* **161**, 567–573.
- YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.
- YAZDI, M. T., WOODWARD, J. R. & RADFORD, A. (1990a). Cellulase production by *Neurospora crassa*: induction and optimisation of the complex. *Enzyme and Microbial Technology* **12**, 116–119.
- YAZDI, M. T., RADFORD, A., KEEN, J. N. & WOODWARD, J. R. (1990b). Cellulase production by *Neurospora crassa*: purification and characterisation of cellulolytic enzymes. *Enzyme and Microbial Technology* **12**, 120–123.