

# Nucleotide sequence of the cellobiohydrolase gene from *Trichoderma viride*

Cheng Cheng, Norihiro Tsukagoshi and Shigezo Udaka\*

Faculty of Agriculture, Nagoya University, Chikusa, Nagoya 464, Japan

Submitted August 9, 1990

EMBL accession no. X53931

The filamentous fungus, *Trichoderma viride*, secretes large amounts of several cellulolytic enzymes (1). We report here the nucleotide sequence of the gene coding for exo-cellobiohydrolase (CBH), which accounts for more than 65% of the total cellulase content. This gene was isolated from a *T. viride* genomic library using a 30mer oligonucleotide based on the sequence of the reported CBHI gene of *T. reesei* (2). The amino acid sequence deduced from the nucleotide sequence is highly homologous (95%) to that of the *T. reesei* CBHI gene. In contrast to the high homology between the enzyme coding regions of the two CBH genes, 67% and 30% homology was found between the 5'-non

coding and 3'-non coding regions, respectively. Southern-blot and Northern-blot analysis showed the presence of at least one CBH gene on the *T. viride* chromosome which was transcribed only at the culture condition that cellulase production was induced.

## REFERENCES

1. Mandels, M. *et al.*, (1975) in *Symposium on Enzymatic Hydrolysis of Cellulose* (Bailey, M. *et al.*, eds), pp. 81–109.
2. Shoemaker, S. *et al.* (1983) *Bio/Technology* 1, 691–696.

```

GTAGCATTGAGGAACCACTGCTTTGGCCAGGAGTAGCGTATGCTGTAGGCAAAGCTCTA 120
GGTGCCACTGCATTGTTGTCGAACATAATGTGATGCTTGGCCAGGCATAATAGCCGCAA 180
AGATAGCCCTCATTGAGCGGAAGTCGGCAACAGTGAAGAGCAGAATATCACATATATA 240
TGGCCAAACCGCGTGTCCCTTCTCCCTTCCCATCTACTCATCAACTCAGATCCCTCC 300
AGAAGACTTGTACATCATCTTTGGGGCATAGCATTCTAGTCCACTAGGACTGGCGCATC 360
ATGATACAAAAGTTGGCCCTCATCTCGGCCCTCTTGGCTACTGCTCGTCTCAGTCCGCC 420
-17  N Y Q K L A L I S A F L A T A R A Q S A
      T G C A C C C T C C A G G C G G A A A C T C A C C C G C C T G A C A T G G C A G A A A T G C T C A T C T G T G C
4   C T I Q A E T H P P L T W Q K C S S G G
ACTTGCACCCAAACAGACAGGCTCCGTGGTCAATCGACCGAACTGGCGCTGGACTACGCC 540
24  T C T Q Q T G S V V I D A N W R W T H A
ACCAACAGCAGCAGCAACTGCTACGACGGCAACTCTGGAGCTCAACCCGTGCCCCTGAC 600
44  T N S S T N C Y D G N A T W S S T L C P D
AATGAGACTTGGCGAAGAAGTGTGCTTGGACGGTCTGCTCCTACCGCTCCACGTACGGA 660
64  N E T C A K N C C L D G A A Y A S T Y G
GTCACCACGAGCGCTGACAGCCTCTCCATTGGCTTCTGCTCACTCAGTCTGCGCAAAGAAC 720
84  V T T S A D S L S I G F V T Q S A Q K N
GTCGGCGCTGCTCTACTTGTAGTGGCGAGTGACACGACCTATCAAGAATCACCCCTGCTT 780
104 V G A R L Y L M A S D T T Y Q E F T L L
GGCAACGAGTCTCTTTTCATGTTGATGTTTCCGACGCTGCCgaagtgaaccaactacact 840
124 G N E F S F D V D V S Q L P
cttgatgccattctctattgattctcagctgactgcttatttaagATGTGGCTTGAAC 900
138 C G L N
GGAGCTCTTTACTTCGTGTCATGGACGGGATGGTGGCGTGACGAAGTATCCCAACCAAC 960
142 G A L Y F V S M D A D G G V T K Y P T N
ACTGCCGTTGCCAAGTACGGCAGCGGCTACTGTGACAGCCAGTCCCTCGTATCTCAAG 1020
162 T A G A K Y G T G Y C D S Q C P R D L K
TTCAACAGCCAGCCCAAGTGTGGGGCTGGGAGCCGCTCTCAACAATGCAAAACAGC 1080
182 F I N G Q A N V E G W E P S N N A N T
GGCATTGGCGGACATGGAAGCTGCTCTGAGATGGATATCTGGGAGGCCAATTCATC 1140
202 G I G G H G S C S C S E M D I W E A N S I
TCTGAGGCTTACTCTCTCATCTTGCACGACCTCGGGCAGGAAATTTGCGAGGGTGAC 1200
222 S E A L T P H P C T T V G Q E I C E G D
TCTTGGCGTGGAACTACTCGGGTGACCGATATGGCGGTACTTGGCCCTGATGGCTGC 1260
S C G G T Y S G D R Y G G T C D P D G C
GATTGGAACCCATATCGCTTGGGCAACACCAGCTTCTATGGCCCCGGCTCCAGCTTCCAG 1320
262 D W N P Y R L G N T S F Y G P G S S F T
CTTGACACCACCAAGAAGTTGACCGTCTGCTCACTGAGTTCGAGACTTGGGTGCCATCAAC 1380
282 L D T T K K L T V V T Q F E T S G A I N
CGATACTATGTCAGAAATGGCGTCACTTCCAGCAGCCCAACCGGAGCTCGGTGATTAC 1440
302 R Y Y V Q N G V T F Q Q P N A E L G D Y
TCTGGCACTCGCTCGAGTACTACTGCGCGGTGAAGAGCGGAGTTCGGCGCTCC 1500
322 S G N S L D D D Y C A A E E A E F G G S
TCTTCTCGGACAAGGCGCCCTTACTCAATTCAAGAAGGCTACTTCCGTTGCCATGGTC 1560
342 S F S D K G G L T Q F K K A T S G G M V
CTGGTCATGAGCCTGTGGGATGACgtgagttccaagaattaacattcaacattgcaacag 1620
362 L V M S L W D D
aatgactagactgatggagagacgatagTACTAGCCCAACATGCTGTGGCTGGACTTAC 1680
370 Y Y A N M L W L D S T
CTACCCGACGGACGAGACCTCCTCCACCCCGGTGCCGTGGAGGCTCCTCCACAG 1740
1740 Y P T D E T S S T P G A V R G S S T S
CTCCGGTGTCTGCTCAGCTCGAGTCCAACCTCCCAACGCCAAGTCGTATACTCCAA 1800
S G V P A Q L E S N S P N A K V V Y S N
CATCAAGTTCGGCCCCACCGGACCCGCAACCTAGCCGCGGAAACCTCCTGGCGG 1860
421 I K F G P I G S T G N P S G G N P P G G
AAACCCCTCCGGCACCACAACACCGCCCGGCTACTCCACTGGAAGCTCCTCCGCGCC 1920
441 N P P G T T T P R P A T S T G S S P G P
TACTCAGCAGCACTATGGCCAGTGGCGTGAATTGGGTACATCGGGCCACCGTCTGCGC 1980
461 T Q T H Y G Q C G G I G Y I G P T V C A
GAGTGGCAGCACTTCCAGGTCTGAAACCCCTACTCTCAGTGTCTGTAAGTACTGT 2040
481 S G S T C Q V L N P Y Y S Q C L ***
GGCAACTCGACATACGTAGTTCGAAAGCTTGGAGTCTGCTGCTTATGGATGAGTTCAT 2100
2100 CTCATTATGGACTAGATGGAGATTTACTTTTGGTATCTACTTCTGAGCTTCCAATAT
ATACGGTATTTCACCTTTGCTGGAATGCTCGCTAGCTTGGCAAGCAGCGCTTCCAGAG 2220
2220 AGACGGACTGATTCTCTGTAACATATGCATTATATAAGACTGAAATAGACCAAAAAAGGA
AAAGTTCCCCACTCGAATATCTGACGGTGTGATTATATGTAGGCCA 2330

```

Nucleotide sequence of CBH. The deduced amino acid sequence is shown below the coding region of the gene and numbered, on the left, from the amino-terminus (Gln) of the mature enzyme. Introns (I and II) are shown in lower-case letters. The putative TATA box sequence is underlined. The amino acid sequence of the amino terminal region of the purified enzyme, which was determined chemically, is double underlined. The stop codon is indicated by asterisks.

\* To whom correspondence should be addressed