Nucleotide sequence of a *Clostridium acetobutylicum* P262 xylanase gene (*xynB*)

Harold Zappe*, Winsome A.Jones and David R.Woods
Department of Microbiology, University of Cape Town, Private Bag Rondebosch, 7700 South Africa

A 1.9 kb *Hind*III-*Eco*RV fragment from pHZ300 (1) was subcloned into the vector pUC19 (2). BAL-31 nuclease was used to generate two overlapping sets of deletions of opposite polarity which were subjected to dideoxy-sequencing (3).

The sequence presented shows the xylanase gene and flanking regions contained within the *Hind*III-*Eco*RV fragment. The open-reading frame encodes for a protein of 262 amino acids (aa). Comparison of the primary protein sequence with other xylanases showed 64% and 40% similarity to a *Bacillus pumilus* xylanase (4) and a *Bacillus subtilis* xylanase (5) respectively. If the relatively non-homologous N-terminal regions of 60 aa (*C. acetobutylicum* xylanase) and 28 aa (*B. pumilus* xylanase) were excluded in this comparison, then their similarity was increased to 71%. The endoglucanase (6) and *glnA* (7) genes of *C. acetobutylicum* also showed homology to the corresponding genes from *Bacillus* spp.

A putative signal peptide of 28 aa (underlined) was identified,

but did not appear to function in *Escherichia coli* (1). The 3' flanking region contained an imperfect pair of inverted repeats (arrows) which could be implicated in transcription termination (8). The characteristics of this gene product (1) were similar to the *C. acetobutylicum* xylanase B (*xynB*) described by Lee *et al.* (9).

REFERENCES

1. Zappe *et al.* (1987) *Appl. Microbiol. Biotechnol.* **27**, 57−63.
2. Yanisch-Perron *et al.* (1985) *Gene* **33**, 103−119.
3. Sanger *et al.* (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463−5467.
4. Fukusaki,E. *et al.* (1984) *FEBS Lett.* **171**, 197−201.
5. Paice,M. *et al.* (1986) *Arch. Microbiol.* **144**, 201−206.
6. Zappe,H. *et al.* (1988) *Appl. Environ. Microbiol.* **54**, 1289−1992.
7. Janssen,P.J. *et al.* (1988) *J. Bacteriol.* **170**, 400−408.
8. Platt,T. (1981) *Cell* **24**, 10−23.
9. Lee,S.F. *et al.* (1987) *Appl. Environ. Microbiol.* **53**, 644−650.

```
GCAGAACTTGAAAAATAAATAAAGAGATACTAAATAAAAATCATTAGAATTTAAAGTGTATAGTTAAGGTTAAATATTAA      80
AAAACATAAGAACAAAGATAGAGTGGGGGATTAAATCACTGATGTTATTATTTGCTTTTTATGCGTTATTTTGGACGTTT     160
CGGGATAAATAATACTATTTAATGTTTATATATGGTAATATGTAAGAGAAATTGTGTTAATCGTTTTCTGTTAGAATTTG     240

GAGGGTATAACTCCATAAACGATTAATCTAAATTTTAGTAAACTAGGAGGGGAATTAAATATGTTGAGAAGAAAAGTAAT     320
                                                      M  L  R  R  K  V  I
TTTTACTGTTTTAGCAACATTAGTTATGACAAGTTTAACTATAGTTGACAATACTGCTTTTGCAGCAACAAATTTAAACA     400
 F  T  V  L  A  T  L  V  M  T  S  L  T  I  V  D  N  T  A  F  A  A  T  N  L  N  T
CAACAGAAAGTACTTTTAGTAAAGAAGTTTTAAGTACGCAAAAAACTTATTCAGCATTTAACACACAGGCAGCACCAAAA     480
   T  E  S  T  F  S  K  E  V  L  S  T  Q  K  T  Y  S  A  F  N  T  Q  A  A  P  K
ACGATTACCTCAAATGAAATTGGTGTTAATGGCGGCTACGACTATGAACTTTGGAAGGACTATGGAAACACCAGTATGAC     560
 T  I  T  S  N  E  I  G  V  N  G  G  Y  D  Y  E  L  W  K  D  Y  G  N  T  S  M  T
GTTGAAGAATGGTGGCGCCGTTTAGTTGTCAATGGAGTAATATCGGAAATGCATTATTCCGTAAAGGCAAGAAATTCAATG     640
 L  K  N  G  G  A  F  S  C  Q  W  S  N  I  G  N  A  L  F  R  K  G  K  K  F  N  D
ATACCCAGACATACAAACAGCTTGGAAATATATCAGTAAACTATGATTGCAATTATCAGCCATATGGCAATTCCTATTTG     720
   T  Q  T  Y  K  Q  L  G  N  I  S  V  N  Y  D  C  N  Y  Q  P  Y  G  N  S  Y  L
TGTGTATATGGATGGACGAGTAGTCCTCTTGTAGAATATTATATAGTTGATAGCTGGGGCAGCTGGAGACCACCTGGCGG     800
 C  V  Y  G  W  T  S  S  P  L  V  E  Y  Y  I  V  D  S  W  G  S  W  R  P  P  G  G
AACATCAAAGGGTACAATTACAGTCGATGGTGGTATCTATGACATATATGAGACCACTCGAATCAACCAGCCTTCAATTC     880
   T  S  K  G  T  I  T  V  D  G  G  I  Y  D  I  Y  E  T  T  R  I  N  Q  P  S  I  Q
AAGGCAATACAACTTTTAAGCAGTATTGGAGTGTCCGCAGAACTAAACGAACTAGCGGAACAATATCTGTCAGCAAACAC     960
   G  N  T  T  F  K  Q  Y  W  S  V  R  R  T  K  R  T  S  G  T  I  S  V  S  K  H
TTTGCCGCTTGGGAAAGTAAAGGAATGCCACTTGGAAAAATGCATGAAACTGCATTTAACATAGAGGGGTACCAAAGCAG    1040
 F  A  A  W  E  S  K  G  M  P  L  G  K  M  H  E  T  A  F  N  I  E  G  Y  Q  S  S
TGGCAAAGCTGATGTGAATAGTATGTCAATTAACATTGGAAAATAAATACCATAAATACTAAAAAAAGTAGCTTTCAAATT    1120
 G  K  A  D  V  N  S  M  S  I  N  I  G  K  *
GAGTTGGAGGCTCCTTTTTTTGAGGTTTTCTAGAATAATCACTATTAATGTTAATAGAAATATCAAAAATAATTCACCTTG    1200

AAATATCAACATTAAAATGGAAACTGAGGAGGTTCACCACATTTATATATTTACGTCGGTTCCATTTTATTGTCTTTGCC    1280
AACTAAAATTATTATTTAACATTAATTGGGTATGTATACAAATTAACAAATAAGTGATAATATTAATTTAAAGATAAATA    1360
TTGGAAAAAATATAAGCTGACATTTTTATAAATAATGTTCGTGAAAGAAAATAATAAATTAATTTTAAGAGCTTAAAGGG    1440
GTGATTATGAGAAA                                                                   1454
```