

Nucleotide sequence of a gene from *Caldocellum saccharolyticum* encoding for exocellulase and endocellulase activity

D.J.Saul, L.C.Williams, D.R.Love, L.W.Chamley and P.L.Bergquist

Department of Cellular and Molecular Biology, University of Auckland, Private Bag, Auckland, New Zealand
Submitted December 7, 1988 Accession no. X13602

Presented is the sequence of the *caIB* gene of *Caldocellum saccharolyticum*; a cellulolytic, obligate anaerobe which grows at temperatures up to 80°C (1). This isolate from a λ library was able to degrade both carboxymethyl cellulose and umbelliferyl-β-D-cellulobiose. The one open reading frame encodes both of these activities. Analysis of deletion mutants showed that MUCCase activity is derived from the N-terminal half of the protein and CMCase from the C-terminus. Highlighted is a putative leader sequence, an *E.coli* like promoter and two 'PT' boxes which divide the protein into three domains. Each of these domains has homology with previously published sequences (2-4). A full description of these data is presented elsewhere.

```

1  CCGTATACA CATAGATTA CTTAGATT CAGTGTACG ATAAAGATA TAGAAGATA GOTTATAAA TTCATAGGA ATTGOTTTT TATGAGGCT
101 CTTGTGTAT GCGAAGTGG CAAATGATA TTAACCTGG ACOTAGAAAT TGTTTCAAAT ATGTTGATG TATTITTTGA ACAGAGGTGG AAMTATTTC
201 AAAGTACGA TTTTAAATT AAAGAGACTG AAGACACAA TGTAAKCTT AAATTTTGA AAGACACCA TTATACAAK AKAAGTTTGT TAANTATGA
301 TACTGTCAA GAANAACAG TAAACGATA TTAACATCA TTAATATCTA TAAATACAA TGTACACAT ATGATATATA TGAATATAC ATAGAAAAC
401 ATGCAAAA GAAATATAA TATTTAAT ATAGOTTAA AAAGCCCAA AAANATATT AGTTTACG TGCACATAT GCTGGAGCC TTTCGCAAC
501 ACCAGACT TTTTAAAG TTTTCTCAT ATTAAGAA TTTTATAC AAATAGCTT TCAMAAATG GCOTCTAAAT TGACAAATTA
601 TATAGTGTG CCAATTTT TTTGACAA ATGAAATA AAANATTA AATAAGAG AAGGGGTTG TGTATAAAT GAGGAAAT CTATTAGGA
701 TTTATCTG AGTTATATA ATAGCOTTA TTCTAGTAT TCTCTGTT GGTGCAATA GTTATTTCC AGTTAAGCT CAGCTGCGC CAGACTGGG
801 TTTCTATG TTATGTAGA GTTAAAGA CAGTTTATG ATAGCOTGG CAATACCTC AAGATATTT ACGATATA TCAACAGC ANTGTATGT
901 I P S L C E S Y K D D F H I G V A I P A R C L S H D T D K H M V L
101 AACTATTTA ACGATATAC ACAGAAAT GAATATAGC CGGAAGACT ATTAGCTGT CAGKCAATA CCGGTTGAG TATGATATT ATACTCTGT
1101 K N H S I L L G L E L A G I E L A G Q T S P G L L S H R F S A D
1201 ASGCTTGT TACTCTGG ATACAAACA AGATAGTAT AGAGCTCAT ACATTAGTT GCATATACA GACACTCAT TGGTTTCTA AGATAGCAA
1301 A F V D F A S T N K I G I R G H T L V W H N Q T P D M W F F K D S N
1401 TGTCAAAAG TTTCCTAAG ATGCATATA ACGAAGATA TTATAGTAT TGTGTAGA TATRAAGATA AAGTGTATC ATGAGCAGT
1501 S D R L L L L K Q L I Y D V W G R Y G K Q V Y A L A M D V
1601 GTCATAGG CTTATCGTA GATACGCA GATAGTATA GACOTCCAC ATGTATAGA AITTTGTCT CTGATGATC TTTATAGCA TTTATAGG
1701 V T N E A I D E H Q P D S Y R R S T W V E I C G P E Y I E K A F I W A
1801 CTCAAGAC AAGCTTAC GCAAGTAT TDCACAGA CTATATAC CTATATAC AAGAAAGA TTTATAGTAT GAGATTTCA ATACTCTGT
1901 H E A D K L Y N D Y N T E I S K R K D F I Y N W V K N L K
201 ACTAAGCT ATACTTAT ATGTATGG TATGCAATG CATATAATG TTAACTGCC ATCTGTTAC GAATATAAA ACATATATA GTTATFATG
2101 S K G I P I H G I G H Q C H I N V N M P S V S E I E H S I F L F S
2201 TGTATACCG GTATGTAGT TCACTATCA GACTGTATA TGAOCTTAT CACTATGCC TCGAGCAA ATATATCC ACACAGCC GATTCTCTC
2301 S I P G L E I H T E L D M S L Y N Y G S S E N S Y T P P Q D L L Q
2401 AAAGACTC TCAGAACTC AAGAGATT TTACATGCT GAAAATATC AAAAATGAT TAAAAGTGT TACTTTTGG GAATAGAG AATATATTC
2501 K Q S Q R Y K E I F T H L K K Y K R V W K S V T M W G L R D W Y
2601 ATGTATAGA TCACTTACG GTAAAGATA TGTGCTGTG TTTTITTTG AAGATATCG TCGAAGCAA GCTTATGGC CATATATGA AGCTTCTGG
2701 W L R S F Y G K N D M P L L F P E D Y S A K P A Y M A V I E A S G
2801 GTACAACT CACTCCAC ACTCAACT ACCCGCAC TACAGTACT ACCAACTA ACACCGAC CAGACCGAC AATACAGG ACTCAACAC
2901 F T S P P F P F P F P F P F P F P F P F P F P F P F P F P F P
301 CACACTTC ACTGTTCAC ACCTCTCA CAGTGTGCA GATAAATA CTATATCA ACAGAGGAC AAGCAGAGC ACGAAGCA TANGCCTG
3101 T P T P V S T P A E T G G Q I K V L Y A N K E T N S T N T I R P W
3201 OTTAAAGTA GTATATGTA GCGAGCTTA CATAGATTG ACGAGGATA CAGTAAATA CTTGTATCAG ATGTATGGT AAAGGACCA AGATGTGCTA
3301 L R V V H S G S S S H D S S F V T F R I V W T V D G R A Q D R A A S V
3401 TCAACTGTC CAGATGAG ACAGACAT GTTACATAA ATTGTGAA OCTAGCCAT ATGTATGTC AGCGGATA TTYCTTAG ATAGATTA
3501 S D M A Q I G A S N V T F K F V K L S S V S G A D Y A L L E I G F K
3601 AAGTGTAG AAGGACTA CAGCTGGA AAGACAGAG AAGATACAG ATAGATTTA ACGAAGATA CTTAGGACT CTTAGCAG GGAATGATG
3701 S G A G Q L Q P F E K D T E I Q T R F H K S D W S H V H Q G N D W
3801 GTCAGATTA CAGAGTGA CAGATTTAG AGAAGTAG AAGTATACG CTTATATGA TGTATGCTG ATGTGGGAC AAGAGCCAG TGGAGCGAC
3901 S M L Q V S T Y G E N E K V T A Y I D G V L V W G Q E P S G A T
401 CTCGCGCA CAGATGACT TGCACCACA GCGACCCCA CACCACACT GAGTCCAGG GTACCGGCA CCGCGCCAG GAGCGGACA AAGATACAA
4101 P D R H E V A F P F P F P F P F P F P F P F P F P F P F P F P
421 GCGAATCT AACTATCA CAGACCA CAGCATAT TCGAATAT ACGAATGAT ATGCGTTA TATATGAT TATAATAT TATATAAGA
4301 P T L T P P P F P P P S S I P D D T H D D W L Y V S G N K I V D K D
4401 GGTGAGCA GTATGTTA CAGTATTA CTGTTTGA TATATACTG GTACAAATG ATTTAGTGT GTATGGTGT GTATCTGAA AGACTCTTT
4501 G R P V W L T G R H H G Y T F G T H V F D G V S Q S N L K D T L
4601 OCTAAATA CATATAGG TTTACTCTG CTAAATCTC AAGACTCAC AAGCTTATA TTGAACTGT CCGAAGAT TTTCCAAA CCAATATAA
4701 A E I A N R G F N L R V P I S A E L I L R W S Q G I Y P K P N I N
4801 ACTACTGT AATCTAG CTTAAGCA AAGACACTG GGAATATT GACATATAG TCGAACATG TCGAACATG TAAAGACTT GHTTAAAA TTTATTTGA
4901 Y Y Y V H E L E G K H L E V F D I V Q T C K E V G L K I M L D
501 TATTCAGT ATAAACAG ACCGATGG ACATATCTA CCGTATGT ATAGAAA AATTACTCA AAGATTTT AATAGGCTG CAAGTGAT
5101 I H S I R T A D A H G H I Y P V W Y D E K F T P E D F Y K C A E M I
5201 ACAATAGT ATAAAGATA TGCACATT ATAGCTTGG ACATAGAGA TACGACACT GGAAAACCT GCAABATAC ACATATCA AATGAGATA
5301 T R V Y K H D D T I A F D L K H E P H G K P W Q D T T F A K W D N
5401 ATCAGACA TATATAT TGAATATG CAGCTAAC AATGCAAAA COTATACTGA ATATAATCC AACTCTTA AZTATATV AAGAATTA
5501 S T D I N H W Y A E T C A K R I L H I N P N L L I V I E G I E
5601 ACCCTACA AAGGACAG TCACTGAC ACAGAACTC TCGAGACT ACTATCAC ATGTCGCG GSAATCCG GAGCTTTAG AAGCTACT
5701 A Y P D D V W T S K S S D Y S T M W G G N L R J V R K Y P
5801 ATATCTGT GTAAATCA AAATAAATT GTATATCC CAGATATA CCGGCCAT CTATTACAG ACCGCTGT TTTACCGG TTCAAAAAG
5901 I N L G K Y Q N K V Y T S P H D Y G P S Y Q G P W Y P G F D Y
6001 ACTTTTAT ACGATGTT TGGCTGTA ATGTCGTA CATATGGA GAAGAAATG CCGCATCT GATGTGGAG TGGGCTGT ACTGTATG
6101 S L L Q D C W P N W A T I M E N I A P L L I G E W G H L D G
6201 ACCGTAAC GAAGATGA TAAATATT ACGAGATT ATCATAGAA ATCATATCA TCACAACT TGTGTTCT ATGCTATA ATGCTACT GGTACACT
6301 A D N E K W H K Y L R D Y I E N H I R H T F W C F H N S G
6401 GAGAGTGG TGTAGATA TTTACAGA TGTAGATA AAATATCT ATTTTGA CCGCTCAT GCGACAGC TAAAGTAG TTTGTGGT
6501 G G L V G D F T T W D E K K Y S F L K P A L W Q D S Q G R F V G L
6601 TGAATACA AAGACCTTA GATTCAAGT GAAAACAT TATATATCA ACATATACA ACATATGA GCTTACAG CTTCCACT CAAATGATA
6701 D H K R F L G E H G K N H H F H E H E F V A H
6801 AATKATTA AATGATAT TATGATATA TGAAGAAAT TTANGTAC TAGGGAGAT TATGTAGTA OCTATNAC AAGAGCTTA AAGAAAG
6901 GAGGTGATA AAAGACTC AAAGATTA GTANTATTA ACGATTAT GTAAAGATA AGTATAGT TCAAACTAT ACCCTCAT ACGATATT
7001 OGGATATT ATCAAACA AAAGAGGGG TTGGGTTAA AAATCTTAA ACTAAAATA GATAAAATA GATTAAGT GGTGTCTCT GTGATTTT
7101 TATGCAAT TTTATTTA TTTGAGGA ACGAAGTCA TGTGTGAAA GCGAGACCA CTTTACTA TGTGTACT CTTCAAAAG CATTATTT
7201 TATGATTC CAGATTTA CACTTCCC ATCATGAC C 4241

```

REFERENCES

1. Donnison, A.M., Brocklesbury, C.M. and Morgan, W.H. (1986) Proc. Int. Congress Microbiol. Manchester. p203
2. O'Neill, G.O., Goh, S.H., Warren, R.A.J., Kilburn, D.G. and Miller Jr., R.C. (1986) Gene 44. 325-330.
3. Mackay, R.M., Lo, A., Willick, G., Zuker, M., Baird, S., Dove, M., Moranelli, F. and Seligy V. (1986) Nucl. Acids Res. 14. 9159-9170.
4. Grépinet, O. and Beguin, P. (1986) Nucl. Acids Res 14: 1791-1799