Note

Nucleotide Sequence of the Xylanase A Gene of Alkalophilic *Bacillus* sp. Strain C-125Tetsuo HAMAMOTO, Hiroshi HONDA,
Toshiaki KUDO and Koki HORIKOSHIThe Riken Institute, Wako-shi,
Saitama 351-01, Japan

Received October 7, 1986

Many alkalophilic microorganisms which grow well at high pH (10.0~11.0) have been isolated in our laboratory. These organisms produce many kinds of enzymes which have optimum pHs in the alkaline region.¹⁾ The genes for certain enzymes isolated from alkalophilic *Bacillus* strains have been cloned in *Escherichia coli*.^{2~5)} Alkalophilic *Bacillus* sp. strain C-125 produces two types of extracellular xylanase: xylanase A, with a very broad pH-activity curve (pH 5 to 11), and xylanase N, with a pH optimum of 7.0.⁶⁾

We cloned the xylanase A gene in *E. coli* with pBR322.^{7,8)} The plasmid, pCX311, thus constructed, contained 2.7 kb and 2.1 kb *Hind*III fragments from alkalophilic *Bacillus* sp. strain C-125.⁷⁾ Xylanase A activity was not found when a 1.6 kb *Cla*I fragment was deleted from pCX311. Therefore, we determined the DNA sequence

around the *Cla*I site in the 2.7 kb fragment and found the xylanase A gene in this fragment. DNA sequencing was done by the dideoxy chain termination method.^{9,10)} The restriction endonuclease map and the strategy for DNA sequencing are shown in Fig. 1. Analysis of the nucleotide sequence of the fragment showed that there was a single open reading frame. A protein of a molecular mass of 45 kDa translated from this open reading frame of 1188 bp (Fig. 2).

The sequence of the N-terminal amino acids of the extracellular form of xylanase A, determined with a peptide sequencer, NH₂-Ala-Gln-Gly-Gly-Pro-Pro-Lys-Ser-Gly-Val-Phe, was identical to that deduced from the DNA sequence (residues 1 to 11). This result suggested that 28 amino acid residues (residues -28 to -1) represented the signal peptide which was removed during secretion of xylanase A. The values obtained from amino acid analysis of extracellular form of xylanase A were consistent with those deduced from the DNA sequence (Table I). The molecular weight of the mature xylanase A as calculated from the DNA sequence was 42479, in good agreement with the molecular weight of 43000 for the extracellular xylanase A produced by *E. coli* carrying plasmid pCX311 coding for xylanase A.⁷⁾

The putative ribosome binding sequence, GAAAGG-AG, observed ten bp upstream of the initiation codon, is complementary to the 3'-end of 16S rRNA of *B. subtilis*, 3'-UCUUUCCUCC-5'.¹¹⁾ S1-nuclease mapping experiment indicated that the mRNA for xylanase A gene was initiated from around 142-bp upstream of the translational initiation site in alkalophilic *Bacillus* sp. strain C-125. The putative promoter region had a consensus sequence for *B. subtilis* σ^{43} factor,¹²⁾ TATAAT (-10) and

FIG 1. Restriction Map of Plasmid pCX311 and Sequencing Strategy.

pCX311 contains two *Hind* III fragments of alkalophilic *Bacillus* DNA in pBR322. The cloned insert of pCX311 is indicated by the open bar, and pBR322 by the dotted bar. Fine mapping of the region surrounding the xylanase A gene and the sequencing strategy (horizontal arrows) are enlarged. The filled and open boxes indicate the coding sequences for the mature protein and the signal peptide, respectively.

HincII
GTTGACGAAAAAGGCATAGAGCATGTATTTAGGACTAA

TCGATCTTTAAGACAACAAAAATAGGACTAAATACAAGTTTAACTACTATAACATTCGGATTATTTTGAAATAACTAGTTTGTTCAC

GGATCAACTAAGTTTATTGTTGCTATAATGGGGCTACAAGGAGCCAGTAGGTAAGATTTAATTTTTCACAACCTTTATGAAAGCCCTT

TCTTTTATCGACAGCGCCACTTTATTGAAAGGAGGTTAGTTAAAGCCGTATTTTTCGATTTTACTTTGTCTACGAAGGAGAAATTTGTG

1 50 S.D. ▽
ATGATTACACTTTT TAGAAAGCCTTTTGTGCTGGGCTAGCGATCTCTTTATTAGTTGGAGGGGGATCGGCAATGTAGCTGCTGCTCAA
MetIleThrLeuPheArgLysProPheValAlaGlyLeuAlaIleSerLeuLeuValGlyGlyGlyIleGlyAsnValAlaAlaAlaGln (1)
(-28) 100 HpaII 150
GGAGGACCACCAAAATCCGAGTCTTTGGAGAAAAAGAAAAAGAAATGATCAGCCTTTTTCGATGGCAAGTTGCTTCTCTTTCTGAGCGA
GlyGlyProProLysSerGlyValPheGlyGluAsnGluLysArgAsnAspGlnProPheAlaTrpGlnValAlaSerLeuSerGluArg

EcoRV 200 250
TATCAAGAGCAGTTTGATATTGGAGCAGCGGTTGAGCCCTATCAATTAGAAGGGAGACAAGCCCAATTTTAAAGCATCATTATAACAGC
TyrGlnGluGlnPheAspIleGlyAlaAlaValGluProTyrGlnLeuGluGlyArgGlnAlaGlnIleLeuLysHisHisTyrAsnSer

300 350
CTTGTGGCGGAAATGCAATGAAGCCTGAATCACTCCAGCCAAGAGAAGGTGAGTGGAACTGGGAAGCGCTGACAAATTTGTGGAGTTT
LeuValAlaGluAsnAlaMetLysProGluSerLeuGlnProArgGluGlyGluTrpAsnTrpGluGlyAlaAspLysIleValGluPhe

400 ClaI 450
GCCCCGAAACATAACATGGAGCTTCGCTTCCACACGCTCGTTTGGCACAGCCCAAGTACCAGAATGGTTTTTCATCGATGAAGACGGCAAT
AlaArgLysHisAsnMetGluLeuArgPheHisThrLeuValTrpHisSerGlnValProGluTrpPheIleAspGluAspGlyAsn

500
CGGATGGTGGATGAACAGACCCAGATAAACGTGAAGCGAATAAACAGCTGTTATTAGAGCGCATGGAACCATATTTAAACCGTTGTT
ArgMetValAspGluThrAspProAspLysArgGluAlaAsnLysGlnLeuLeuLeuGluArgMetGluAsnHisIleLysThrValVal

550 600
GAACGTTATAAAGATGATGTGACTTCATGGGACGTGGTCAATGAAGTTATTGACGATGGCGGGGCGCTGGAATCGGAATGGTATCAA
GluArgTyrLysAspAspValThrSerTrpAspValValAsnGluValIleAspAspGlyGlyGlyLeuArgGluSerGluTrpTyrGln

650 700
ATAACAGGCACCTGACTACATTAAGGTAGCTTTTGAACCGCAAGAAAAATATGGTGGTGAAGAGGCAAGTTGTACATTAATGATTACAAC
IleThrGlyThrAspTyrIleLysValAlaPheGluThrAlaArgLysTyrGlyGlyGluGluAlaLysLeuTyrIleAsnAspTyrAsn

750 800
ACCGAAGTTCCTTCAAAAAGAGATGACCTTTACAACCTGGTGAAAGACTTATTAGAGCAAGGAGTGCCAATTGACGGGGTAGGACATCAG
ThrGluValProSerLysArgAspAspLeuTyrAsnLeuValLysAspLeuLeuGluGlnGlyValProIleAspGlyValGlyHisGln

850 900
TCGCATATCCAAATCGGCTGGCCTTCCATTGAAGATACAAGAGCTTCTTTGAAAAGTTTACGAGTTTAGGATTAGACAATCAAGTAACT
SerHisIleGlnIleGlyTrpProSerIleGluAspThrArgAlaSerPheGluLysPheThrSerLeuGlyLeuAspAsnGlnValThr

950
GAGCTAGACATGAGTCTCTATGGCTGGCCACCGACAGGGGCATACAGTCAATGACGACATTCAGCAGAACTCCTTCAAGCCAGGCA
GluLeuAspMetSerLeuTyrGlyTrpProProThrGlyAlaTyrThrSerTyrAspAspIleProAlaGluLeuLeuGlnAlaGlnAla

1000 1050
GACCGTTACGATCAGCTTTTCGAGTTATACGAAGAATTAGCTGCTGATATTAGCAGTGTAACCTTCTGGGGAATTTGCTGATAACCATACA
AspArgTyrAspGlnLeuPheGluLeuTyrGluGluLeuAlaAlaAspIleSerSerValThrPheTrpGlyIleAlaAspAsnHisThr

1100 1150
TGGCTTGATGGCCGCTAGAGAGTACAATAATGGAGTAGGGATCGATGCACCATTTGTTTTGATCATAACTATCGAGTGAAACCTGCT
TrpLeuAspGlyArgAlaArgGluTyrAsnAsnGlyValGlyIleAspAlaProPheValPheAspHisAsnTyrArgValLysProAla

1191
TACTGGAGAATTATTGATTAGAAGGTAGTACTCAATTAATAGTCTAGCAACGAACAGCTCTGCTTTGGCAGAGCTGTTTTTACTGGA
TyrTrpArgIleIleAsp***
(396)

ACCAAAAGCCTCTATTGTTGAGTAAAAATCATATAATTTCTGAAAGGAACACAAACATTCGACGGTTGCAACTTAAATAACAATGCAGG
TAATGGTAGAAGGTTTTTACCTATGGATAGCGAAAAAGTGAAGAAAAATGATGACAACGTATGAGAGAGGGAGGAAATATTAGTGAGCT
ATGATATCTTAACATTAGCAGCTATCGGCAATCGGATTAATAATGGAGGGGACCTTTTCTGAAATTGATCCCGAATTTTAAAAAATGT
AATTCACAAATGGAGATCGTGCTAGTGAATAGGCATAAATCAATTCCAACATTCAATTAGGCCT

AatI

Fig. 2. The Complete Nucleotide and Amino Acid Sequences of the Alkalophilic *Bacillus* sp. Strain C-125 Xylanase A Gene.

The position of processing of the signal peptide is indicated by an arrowhead. The putative transcriptional start site and ribosomal binding sequence (S.D.) are shown by an arrow and underlining, respectively. The putative promoter consensus sequence is indicated by dashed underlining. The inverted repeat structure for the transcription terminator is indicated by underlining.

TABLE I. AMINO ACID COMPOSITION OF MATURE XYLANASE A

Amino acid	Amino acid analysis	From DNA sequence	Amino acid	Amino acid analysis	From DNA sequence
Lys	17	17	Ala	24	25
His	11	10	Cys	0	0
Arg	20	19	Val	24	24
Asx*	50	50	Met	5	5
Thr	15	15	Ile	18	19
Ser	16	17	Leu	25	26
Glx*	54	55	Tyr	18	18
Pro	16	16	Phe	14	14
Gly	25	26	Trp	ND	12

*Asx represents Asp and Asn; Glx represents Glu and Gln. ND, not determined.

GGATCA (–35). There is one inverted repeat structure downstream of the xylanase A gene followed by thymidine residues, which could play a role in transcription termination.¹³⁾ Similar structures are observed at the downstream of the genes from *Bacillus subtilis*¹⁴⁾ and an alkalophilic *Bacillus* strain.³⁾

Acknowledgment. We thank Dr. R. H. Doi and Dr. K. N. Joblin for reading the manuscript and for many useful discussions. This work was partially supported by a Grant for Life Science Promotion from the Riken Institute.

REFERENCES

- 1) K. Horikoshi and T. Akiba, "Alkalophilic Microorganisms, a New Microbial World," Springer-Verlag, New York, 1982.
- 2) C. Kato, T. Kudo, K. Watanabe and K. Horikoshi, *Eur. J. Appl. Microbiol. Biotechnol.*, **18**, 339 (1983).
- 3) C. Kato, T. Kudo, K. Watanabe and K. Horikoshi, *J. Gen. Microbiol.*, **131**, 3317 (1985).
- 4) T. Kudo, A. Ohkoshi and K. Horikoshi, *J. Gen. Microbiol.*, **131**, 2825 (1985).
- 5) N. Sashihara, T. Kudo and K. Horikoshi, *J. Bacteriol.*, **158**, 503 (1984).
- 6) H. Honda, T. Kudo, Y. Ikura and K. Horikoshi, *Can. J. Microbiol.*, **31**, 538 (1985).
- 7) H. Honda, T. Kudo and K. Horikoshi, *J. Bacteriol.*, **161**, 784 (1985).
- 8) H. Honda, T. Kudo and K. Horikoshi, *Agric. Biol. Chem.*, **49**, 3165 (1985).
- 9) F. Sanger, S. Nicklen and A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5463 (1977).
- 10) J. Messing, "Third Cleveland Symposium on Macromolecules: Recombinant DNA," Elsevier, Amsterdam, 1981, p. 143.
- 11) J. McLaughlin, C. Murray and J. C. Rabinowitz, *J. Biol. Chem.*, **256**, 11283 (1981).
- 12) W. C. Johnson, C. P. Moran, Jr. and R. Losick, *Nature*, **302**, 800 (1983).
- 13) M. Rosenberg and D. Court, *Ann. Rev. Genet.*, **13**, 319 (1979).
- 14) M. A. Gitt, L. Wang and R. H. Doi, *J. Biol. Chem.*, **260**, 7178 (1985).