The complete nucleotide sequence of the xylanase gene (*xynA*) of *Bacillus pumilus*

Eiichiro Fukusaki, Watanalai Panbangred, Atsuhiko Shinmyo and Hirosuke Okada*

Department of Fermentation Technology, Osaka University, Yamada-oka Suita-shi, Osaka 565, Japan

Received 19 March 1984

The complete nucleotide sequence of the xylanase (EC 3.2.1.8) gene (*xynA*) and its flanking regions of *Bacillus pumilus* IPO, a hyperproducer of xylanase, was determined. A 684 bp open reading frame for xylanase gene was observed. The amino acid sequence of the N-terminal region of xylanase was determined to be Arg-Thr-Ile-Thr-, suggesting the processing at Ala^{27} of pre-xylanase. The amino acid composition and M_r (22384) of xylanase deduced from DNA sequence agreed with the results obtained with the purified enzyme. The signal sequence consisted of 27 amino acids, of which 3 were basic amino acid residues in the region near the N-terminus and 15 were hydrophobic amino acid residues. The ribosome binding sequence complementary to the 3'-end of 16 S rRNA of *B. subtilis* was found 7 bp upstream of the initiation codon, ATG.

Bacillus pumilus *Xylanase* *DNA sequence* *Signal sequence* *Gene cloning* *Ribosome binding sequence*

1. INTRODUCTION

Xylanase is a potentially important enzyme for the utilization of agricultural wastes, and its secretion provides a suitable example for the study of the secretion mechanism of bacterial enzymes. Though the genus *Bacillus* produces a variety of extracellular enzymes, the secretion signals of too few enzymes have been analyzed to identify the consensus signal peptide sequence.

In [9,10] we reported the cloning of the xylan degradation genes, the xylanase (EC 3.2.1.8, 1,4-β-D-xylan xylanohydrolase) gene (*xynA*) and the β-xylosidase (EC 3.2.1.37, 1,4-β-D-xylan xylohydrolase) gene (*xynB*), of *B. pumilus* IPO in *E. coli*. Though xylanase is an extracellular enzyme in *B. pumilus* IPO, it was not secreted from *E. coli* cells harboring a hybrid plasmid encoding the *xynA* gene but was accumulated in the cytoplasm. This failure of secretion was thought to be caused either by partial deletion of the secretion signal or by the miss-recognition by *E. coli* of the signal useful for the genus *Bacillus*. Here, we describe the complete sequence of the *xynA* gene including the ribosome binding site, secretion signal and structural regions.

2. MATERIALS AND METHODS

2.1. *Bacterial strains and plasmids*

B. pumilus IPO, a hyper-producer of xylanase, and *E. coli* C600 (*leu-6, thr-1, thi-1, lacY1, supE44, tonA21, hsdR, hsdM*) were described in [9]. Plasmids pOXN29, consisting of pBR322 and a 14.4 kilo base pair (kbp) *B. pumilus* DNA fragment coding the xylanase gene (*xynA*) and the β-xylosidase gene (*xynB*), and pOXN391R, constructed by inserting the 6.0 kbp *Bgl*II fragment of pOXN29 containing *xynA* at the *Bam*HI site of pBR322, have been described [9].

2.2. *DNA sequencing*

Single end labeled DNA fragments were isolated and sequenced by the chemical modification

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

methods in [3]. Degradation products were resolved on 20 or 8% polyacrylamide sequencing gel.

2.3. *N-terminal amino acids sequencing of* B. pumilus *xylanase*

Purified *B. pumilus* xylanase was sequentially degraded as in [1]. PTH-amino acids were identified with an HPLC system (PN101, Yanaco Co., Kyoto, Zorbox CN analytical column, Du Pont Co., DE).

3. RESULTS

3.1. *Sequencing strategy*

The structure of the plasmid pOXN391R is shown in fig.1, in which the location of *xynA* gene was restricted between the *Eco*RI site at 2.1 kbp and the *Hin*dIII site at 3.2 kbp on the map as in [9]. A deletion plasmid pOXN392R, in which the 3.2 kbp *Hin*dIII fragment (3.2–6.4 kbp region on the map) was deleted from pOXN391R, was constructed by partial digestion with *Hin*dIII and religation. *E. coli* transformants harboring pOXN392R were Ap^r, Tc^s and xylanase-positive.

The nucleotide sequence from the *Eco*RI site at 2.1 kbp to the *Hin*dIII site at 3.2 kbp was determined by the strategy outlined in fig.1. The nucleotide sequence for the coding region of xylanase has been determined from both strands.

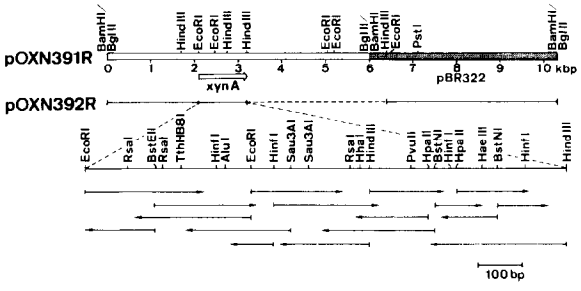

Fig.1. Structure of plasmids pOXN391R and pOXN392R, and strategy of nucleotide sequencing. pOXN391R is a hybrid plasmid consisting of the 6 kbp *Bgl*II fragment of *B. pumilus* DNA (open bar) and pBR322 (dotted bar). The maximum range of the location of the *xynA* gene and the direction of transcription determined in [9] are shown by the open arrow. pOXN392R deleted the 3.2 kbp *Hin*dIII fragment shown by a broken line from pOXN391R. Fine mapping of the regions containing the *xynA* gene and sequencing strategy are enlarged.

3.2. *Nucleotide sequence of* xynA

The nucleotide sequence of 1070 bp covering the entire *xynA* gene and its flanking regions is shown in fig.2. The sequence is consistent with the observed restriction fragments. Of three reading frames found, one was a 684 bp open reading frame beginning at 61 bp. The amino acid sequence of the N-terminal region of *B. pumilus* xylanase which is excreted in the culture medium, was determined to be Arg-Thr-Ile-Thr- by sequential Edman degradation followed by identification of the PTH-amino acids (not shown). This finding suggests that the signal peptide consisting of 27 amino acid residues is processed between Ala^{27} and Arg^{28}. The processed xylanase was deduced to consist of 201 amino acid residues, corresponding to an M_r of 22384, which agrees with the M_r of purified xylanase of *B. pumilus* estimated by the methods of equilibrium ultracentrifugation (20000) and SDS-polyacrylamide gel electrophoresis (SDS–PAGE) (24000) [8]. The number of residues of each amino acid contained in the xylanase molecule estimated from the nucleotide sequence agreed with our previous analysis of the amino acid composition of the enzyme, except for tryptophan which might be overestimated [8]. The ribosome binding sequence complementary to the 3′-end of 16 S rRNA of *B. subtilis*, 3′-UCUUUCCUCCACUAG-5′ [4], was observed 7 bp upstream of the initiation codon, ATG.

3.3. *Signal sequence*

The amino acid sequence of the signal peptide of pre-xylanase was deduced from the DNA sequence. It consists of 27 amino acids, of which 3 are basic amino acid residues, Arg and Lys, concentrated in the region near the N-terminus, and 15 are hydrophobic amino acid residues, Leu, Phe, Val, Ile, and Ala.

3.4. *Codon usage*

Codon utilization for pre-xylanase is presented in table 1. There appears to be a bias for codon selection, for example, UUU for Phe, GAU for Asp, GAA for Glu, CCA for Pro, and AGA and

```
                    20                  40                  60                  75                  90
  GAATTCATTT CATCTTAGAG ATGACAGAAT TAAAAGGATG AAAAAGGAGA GGAATGACGA ATG AAT TTG AGA AAA TTA AGA CTG TTG TTT GTG ATG
                                                                    MET ASN LEU ARG LYS LEU ARG LEU LEU PHE VAL MET

       105                 120                 135                 150                 165                 180
TGT ATT GGA CTG ACG CTT ATA CTG ACG GCT GTA CCA GCC CAT GCG AGA ACC ATT ACG AAT AAT GAA ATG GGT AAC CAT AGC GGG TAC
CYS ILE GLY LEU THR LEU ILE LEU THR ALA VAL PRO ALA HIS ALA Arg Thr Ile Thr Asn Asn Glu Met Gly Asn His Ser Gly Tyr

           195                 210                 225                 240                 255                 270
GAT TAT GAA TTA TGG AAG GAT TAT GGA AAC ACC TCG ATG ACA CTC AAT AAC GGC GGG GCA TTT AGT GCA GGC TGG AAC AAT ATC GGA
Asp Tyr Glu Leu Trp Lys Asp Tyr Gly Asn Thr Ser Met Thr Leu Asn Asn Gly Gly Ala Phe Ser Ala Gly Trp Asn Asn Ile Gly

               285                 300                 315                 330                 345
AAT GCT TTA TTT AGA AAA GGG AAA AAG TTT GAT TCC ACT AGA ACT CAC CAT CAG CTT GGC AAC ATA TCC ATC AAT TAC AAC GCA AGT
Asn Ala Leu Phe Arg Lys Gly Lys Lys Phe Asp Ser Thr Arg Thr His His Gln Leu Gly Asn Ile Ser Ile Asn Tyr Asn Ala Ser

360                 375                 390                 405                 420                 435
TTT AAC CCA AGC GGG AAT TCC TAT CTA TGT GTC TAT GGC TGG ACA CAA TCT CCA TTA GCA GAA TAC TAC ATT GTT GAT TCA TGG GGC
Phe Asn Pro Ser Gly Asn Ser Tyr Leu Cys Val Tyr Gly Trp Thr Gln Ser Pro Leu Ala Glu Tyr Tyr Ile Val Asp Ser Trp Gly

    450                 465                 480                 495                 510                 525
ACA TAT CGT CCA ACA GGA GCG TAT AAA GGA TCA TTT TAT GCT GAT GGA GGC ACA TAT GAC ATT TAT GAA ACA ACC CGT GTC AAT CAG
Thr Tyr Arg Pro Thr Gly Ala Tyr Lys Gly Ser Phe Tyr Ala Asp Gly Gly Thr Tyr Asp Ile Tyr Glu Thr Thr Arg Val Asn Gln

       540                 555                 570                 585                 600                 615
CCT TCC ATT ATC GGG ATC GCA ACC TTC AAG CAA TAT TGG AGT GTA CGT CAA ACG AAA CGT ACA AGC GGA ACG GTC TCC GTC AGC GCG
Pro Ser Ile Ile Gly Ile Ala Thr Phe Lys Gln Tyr Trp Ser Val Arg Gln Thr Lys Arg Thr Ser Gly Thr Val Ser Val Ser Ala

           630                 645                 660                 675                 690                 705
CAT TTT AGA AAA TGG GAA AGC TTA GGG ATG CCA ATG GGG AAA ATG TAT GAA ACG GCA TTT ACT GTA GAA GGC TAC CAA AGC AGC GGA
His Phe Arg Lys Trp Glu Ser Leu Gly Met Pro Met Gly Lys Met Tyr Glu Thr Ala Phe Thr Val Glu Gly Tyr Gln Ser Ser Gly

               720                 735                 760                 780                 800
AGT GCA AAT GTG ATG ACC AAT CAG CTG TTT ATT GGC AAC TAAAAA AGTCAAAGAA AAGAGCCGGG AGCAAAACTC CTGGCTTTTT CTATCATAAT T
Ser Ala Asn Val Met Thr Asn Gln Leu Phe Ile Gly Asn

                   820                 840                 860                 880                 900
TTTCAACTT CGACTCTGCC GGGAAAGAAC GTTCCGGAAA AGAACGTCGC ACCGCCGCCC ATATCTGCCA AGCGATCAGG TGTGAGGCCA TTCACCAAAT GTTTTT

           920                 940                 960                 980                 1000
TGCC TTTTTTGGTC TGCCCATAAT CCTGGCTGAC AACAACACCA GATAACACAT TTTGTCCGAC TGACACGATC AGCTCGCATT CTCCTCGATG ATTCAGCACA

      1020                1040                1060
CGAACCATGT CCCCATCTTC AATCTGTCTT TCTTTTGCAT CTTGTTCATT CATATGAGGC
```

Fig.2. The complete nucleotide sequence and amino acid sequence for xylanase. The 27 amino acid signal peptide is shown in capital letters. The position of processing of pre-xylanase is shown by the arrow. The ribosome binding sequence complementary with the consensus sequence [4] is underlined, and the possible promoter sequences are shown by broken underlining (see text).

CGU for Arg. There are 24 Gly, but GGU is used only once.

4. DISCUSSION

We have constructed a hybrid plasmid, pOXN29 coding both the xylanase and β-xylosidase genes of *B. pumilus* IPO [9,10]. Here, the complete nucleotide sequence of the xylanase gene (*xynA*) and its flanking regions was determined. It was found that *xynA* consisted of 684 bp, including an 81 bp stretch for the signal peptide at the 5′-end. These results are consistent with the amino terminal sequence and amino acid composition of xylanase synthesized in vivo. The ribosome binding sequence was found 7 bp upstream of the translation initiation codon. Candidates for promoter sequences, the -35 and -10 regions, were found 48 bp upstream (TTTCAT) and 26 bp upstream (TTAAAA) of the initiation codon, ATG, respectively. However, they are different from the consensus sequences, TTGACA and TATAAT, of *B. subtilis* [5], and the possibility remains that the true promoter is located upstream of the 5′-end determined here.

Table 1

Codon usage of pre-xylanase

Amino acid	Codon	No. of codons used	Amino acid	Codon	No. of codons used
Phe	UUU	9	Ser	UCU	1
	UUC	1		UCC	5
Leu	UUA	5		UCA	2
	UUG	2		UCG	1
	CUU	2	Pro	CCU	1
	CUC	1		CCC	0
	CUA	1		CCA	5
	CUG	4		CCG	0
Ile	AUU	6	Thr	ACU	3
	AUC	4		ACC	5
	AUA	2		ACA	7
Met	AUG	8		ACG	6
Val	GUU	1	Ala	GCU	3
	GUC	4		GCC	1
	GUA	3		GCA	7
	GUG	2		GCG	3
Tyr	UAU	11	Cys	UGU	2
	UAC	5		UGC	0
His	CAU	4	Trp	UGG	6
	CAC	1	Arg	CGU	4
Gln	CAA	4		CGC	0
	CAG	3		CGA	0
Asn	AAU	11		CGG	0
	AAC	8	Ser	AGU	4
Lys	AAA	7		AGC	7
	AAG	3	Arg	AGA	6
Asp	GAU	5		AGG	0
	GAC	1	Gly	GGU	1
Glu	GAA	7		GGC	8
	GAG	0		GGA	8
				GGG	7

Authors in [2] reported that the chloramphenicol acetyltransferase gene of *B. pumilus* started translation by TTG instead of ATG. Such a novel initiation codon was also found in *Staphylococcus aureus* β-lactamase [4]. In the nucleotide sequence determined here, no open reading frame started by TTG was found except those at 67 bp and 85 bp in the same frame of *xynA*. These codons and ATG at 94 bp might be excluded from the functional initiation codon, because of the long distance from the ribosome binding site or unusual N-terminal amino acid sequence of the signal peptide.

The signal sequence of pre-xylanase deduced from the DNA sequence, the presence of basic amino acids in the N-terminal region, the abundance of hydrophobic amino acids, and the processing at the carboxyl residue of Ala^{27} in pre-xylanase, are consistent with the features of secreted enzymes of Gram-positive bacteria, such as *B. subtilis* α-amylase [11], *B. amyloliquefaciens* α-amylase [7], *B. licheniformis* penicillinase [6] and *S. aureus* β-lactamase [4]. Recently, we have subcloned the *xynA* gene and its flanking regions in *B. subtilis* using pUB110 as cloning vector, and found xylanase to be excreted into the culture medium (unpublished). On the other hand, *E. coli* C600 cells harboring pOXN29 or pOXN391, in which *B. pumilus* DNA was ligated with pBR322 in reverse orientation to pOXN391R, accumulated xylanase in the cytoplasm [9]. The M_r of xylanase prepared from *E. coli* cells harboring pOXN391 was identical to that of *B. pumilus* in SDS–PAGE (unpublished). These results suggest that *E. coli* cells failed to transport the processed xylanase to the cytoplasmic membrane or periplasm. The detailed structure of the N-terminus of the xylanase located in the cytoplasm in *E. coli* (pOXN391) is under study.

ACKNOWLEDGEMENTS

We are grateful to Professor F. Sakiyama, Institute for Protein Research, Osaka University for the partial amino acid sequencing of xylanase, and to N. Kanzaki for technical help. This work was partly supported by Grants-in-Aid for Scientific Research 58040051 from the Ministry of Education, Culture and Science, Japan.

REFERENCES

[1] Allen, G. (1981) Laboratory Techniques in Biology and Molecular Biology: Sequencing of Proteins and Peptides (Work, T.S. and Burdon, R.H. eds) pp.161–236, Elsevier, Amsterdam, New York.
[2] Harwood, C.R., Williams, D.M. and Lovett, P.S. (1983) Gene 24, 163–169.
[3] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[4] McLaughlin, J.R., Murray, C.L. and Rabinowitz, J.C. (1981) J. Biol. Chem. 256, 11283–11291.

[5] Moran, C.P. jr, Lang, N., LeGrice, S.F.J., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J. and Losick, R. (1982) Mol. Gen. Genet. 186, 339–346.
[6] Neugebauer, K., Sprengel, R. and Schaller, H. (1981) Nucleic Acids Res. 9, 2577–2588.
[7] Palva, I., Pettersson, F., Kalkkinen, N., Lehtovaara, P., Sarvas, M., Sörderlund, H., Takkinen, K. and Kääriäinen, L. (1981) Gene 15, 43–51.
[8] Panbangred, W., Shinmyo, A., Kinoshita, S. and Okada, H. (1983) Agric. Biol. Chem. 47, 957–963.
[9] Panbangred, W., Kondo, T., Negoro, S., Shinmyo, A. and Okada, H. (1983) Mol. Gen. Genet. 192, 335–341.
[10] Panbangred, W., Kawaguchi, O., Tomita, T., Shinmyo, A. and Okada, H. (1984) Eur. J. Biochem., in press.
[11] Yamazaki, H., Ohmura, K., Nakayama, A., Takeichi, Y., Otozai, K., Yamasaki, M., Tamura, G. and Yamane, K. (1983) J. Bacteriol. 156, 327–337.