Vol. 176, No. 1, 1991
April 15, 1991

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
Pages 321-327

Nucleotide sequence of *pnl* gene from *Erwinia carotovora* Er

Hiroki Ohnishi, Tamotsu Nishida, Aruto Yoshida, Yoshiyuki Kamio and Kazuo Izaki

Department of Agricultural Chemistry, Faculty of Agriculture
Tohoku University, Amamiya-machi, Sendai, 981 Japan

Received March 1, 1991

SUMMARY: The nucleotide sequence of *pnl* gene encoding pectin lyase (PNL; EC4.2.2.10)from *Erwinia carotovora* Er was determined. The structural gene of *pnl* consisted of 942 base pairs. An open reading frame that could encode a 33,700 dalton polypeptide consisting 314 amino acids was assigned. The molecular size of the polypeptide predicted from the amino acid composition was close to the value of PNL determined in *E.carotovora* Er. The nucleotide sequence of the 5'-flanking region showed the presence of the consensus sequence of ribosome binding site, Pribnow box and the RNA polymerase recognition site in *E.carotovora* and *Escherichia coli*. Between the presumed Pribnow box and the ribosome binding site, two pairs of inverted repeats were found. By comparing the predicted amino acid sequences of *pnl*, several reported bacterial pectate lyases and *Aspergillus niger* pectin lyase, short regions of homology were found despite the different substrate specificities of these enzymes.

Many phytopathogenic *Erwinia* species including *Erwinia carotovora* Er secrete the pectolytic enzymes such as pectate lyase (PL; EC 4.2.2.2) which are thought to be one of the enzymes causing the soft-rot. In addition to PL, some strains of *E.carotovora* produce pectin lyase (PNL; EC 4.2.2.10) in response to DNA-damaging agents such as nalidixic acid, mitomycin C or UV light(1). In most strains, PNL production is accompanied by cell lysis and production of a bacteriocin (2). In addition, PNL production in *E.carotovora* subsp. *carotovora* requires a functional *recA* gene (3). We have recently cloned the *pnl* gene from *E.carotovora* Er and expressed the gene by using *tac* promoter in *Escherichia coli* (4). In the present study, we determined the complete nucleotide sequence of the *pnl* gene and its flanking region of *E.cartovora*

0006-291X/91 $1.50

Er to study the mechanism of the expression of *pnl* gene and the chemical structure of PNL.

MATERIALS AND METHODS

Bacterial strains and plasmids. *Escherichia coli* JM109 {*recA1*, λ^-, Δ(*lac-proAB*), *endA*1, *gyrA*96, *thi*, *hsdR*17, *relA*1, *supE*44, [F',*traD*36, *proAB*,*lacI*^{q} ZΔ*M*15]} was used as a host strain for recombinant plasmids. Plasmid pUC118 and pUC119 were used as cloning vectors. *E.coli* MV1184 {*ara*,Δ(*lac-pro*), *strA*, *thi*,(ϕ80 *lacZ*Δ*M*15), Δ(*srl-recA*)306::Tn10(tet^r), [F',*traD*36, *proAB*, *lacI*^{q} Z-Δ*M*15]}, and M13KO7 were used as a host and a helper phage, respectively for preparation of single stranded DNA.

DNA sequencing. The 2.1 kb of *Stu*I-*Eco*RI fragment of pTN2159 (4) was subcloned into *Sma*I site of pUC119 and pUC118 after *Eco*RI site of the fragment was treated with S1 nuclease. A series of deletion derivatives of each subclones were obtained by exonuclease III and mung bean nuclease digestion, according to the procedures described by Henikoff (5). DNA sequencing was performed by the dideoxy chain termination method of Sanger *et al.* (6).

RESULTS AND DISCUSSION

Nucleotide sequence of the *pnl* gene from pTN2159. The *Eco*RI-*Stu*I fragment in pTN2159 contains the *pnl* gene of *E.cartovora* Er (4). We have determined the nucleotide sequence of the fragment in which contained the *pnl* gene and its 5'- and 3'-flanking regions following the sequencing strategy shown in Fig. 1. The nucleotide sequence of the fragment was comprising 1,286 bp as shown in Fig. 2. Within this sequence, we can identify an open reading frame which begins with an ATG codon at position 290 and terminates with TAA codon at position 1232. The amino acid se-

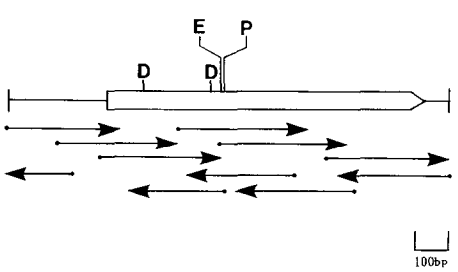

Fig.1. **Physical map and sequence strategy of the 1286 bp *Stu*I-*Eco*RI fragment.** Lines and arrows indicate the direction and extent of a portion of DNA fragment of which the nucleotide sequence have been determined and are aligned in the 5' to 3' direction. Restriction sites: D, *Dra*I; E, *Eco*RV; P, *Pvu*II.

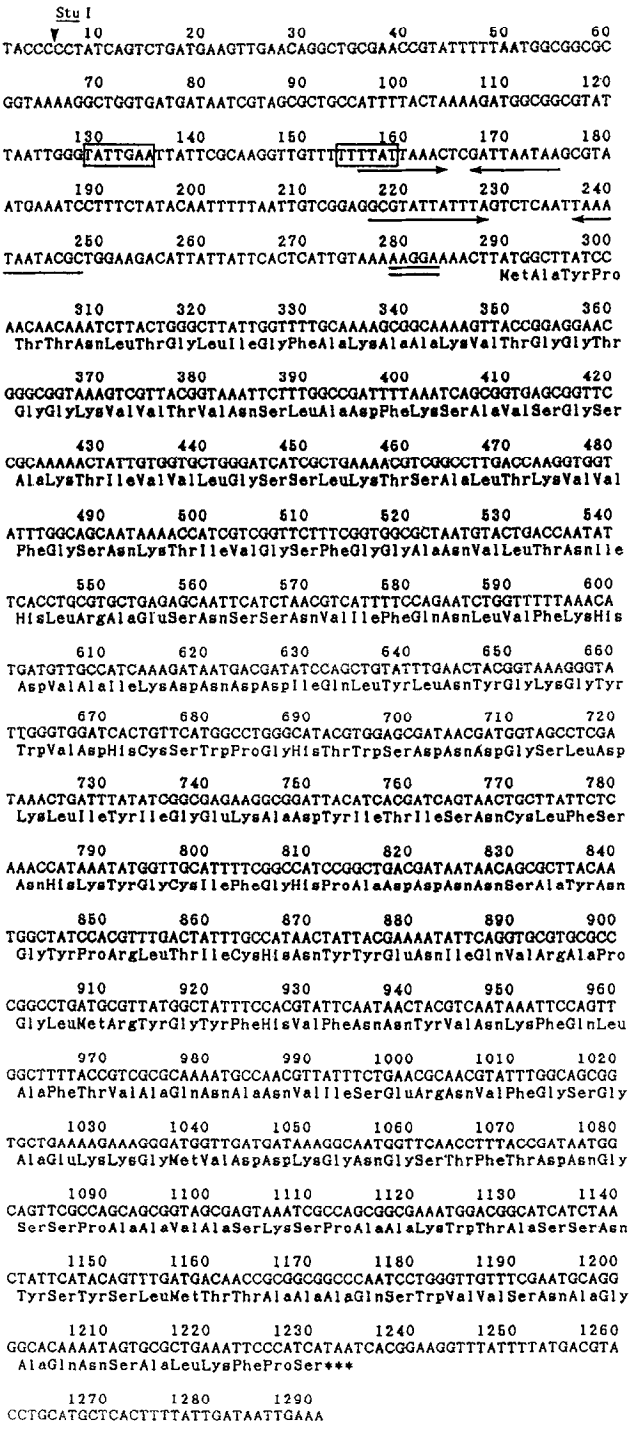

Fig.2. **Nucleotide sequence of the *StuI-EcoRI* fragment and the deduced amino acid sequence of pectin lyase.** The sequence is arranged so that bp 6 is the first nucleotide in the recognition site for *StuI*. The strand shown is the 5' to 3' direction. The deduced amino acid sequence is given bellow the corresponding nucleotide sequence. The predicted ribosomal binding site and promoter site are indicated by double underlines and rectangular boxes, respectively. Arrows represent inverted repeats.

quence of PNL deduced from the DNA sequence corresponds to a polypeptide with an apparent molecular weight of 33.7 kDa, comprising 314 amino acid residues, which coincides with that determined by SDS-PAGE (4,7). The first 18 amino acid residues deduced from the nucleotide sequence exactly corresponded with the N-terminal ones of the purified PNL (4). There was no indication of processing of N-terminus. The sequence of PNL was analysed for hydropathy and overall polarity according to the methods of Kyte and Doolittle (8) and Capaldi and Vanderkooi (9), respectively. The polarity of PNL was 43.0% and hydropathy analysis revealed no significant hydrophobic segment in PNL.

Codon usage. The codon usage for the *pnl* gene was computed, and the following observations were made concerning the strong bias in the usage. (a) Codon AAA is used by 18 out of 21 Lys residues and GAT is used by 13 out of 15 Asp residues. (b) Neither AGA nor AGG is used for Arg at all in the gene. In addition, although 15 Ile residues are present in the protein, none of them used codon ATA. These characteristics in the codon usage is very similar to those of codon used in PLI and PLIII (10,11) except in the case of Asp.

Amino acid homology among the PNL, PL from *Erwinia* species and fungal PNL. The nucleotide sequence and deduced amino acid sequence of the PLI (10) and PLIII (11) of *E.carotovora* Er and PLa (12) and PLb (13,14) from *E.carotovora* subsp. *carotovora* were determined. Recently, Gysler *et al.* cloned the *pelD* gene from *Aspergillus niger*, coding for a pectin lyase D (PLD) and reported the sequence and the predicted amino acid sequence of the gene (15). Amino acid homology between bacterial pectate lyase (PL), our PNL and *A.niger* PLD, shown in Table 1, revealed several regions of homology among them. Although the homologous regions are relatively short, they aligned well. It should be noted that

Table 1. **Comparison of deduced amino acid sequences of *E. carotovora* Er PNL, *Aspergillus niger* PLD and *Erwinia* species pectate lyases**

Protein [a]		Region [b]	Sequence [c]
E.carot. Er	PNL	I (15)	AKAAKVTGGTGGKVV
A.niger	PLD	I (15)	ARAAAV-GVSGTPVG
E.carot. Er	PL I	I (54)	IEAAKKDSS-GKAVK
E.carot. Er	PL III	I (54)	IEAAKLDSN-GKKVK
E.carot. EC	PLa	I (54)	IEEAQLDSK-GKKLK
E.carot. EC	PLb	I (54)	IEAAKVDSK-GKKVK
E.carot. Er	PNL	II (45)	AKTIVVLGSSLKTSALTK
A.niger	PLD	II (59)	AR-VIVL--S-KTFDFTD
E.carot. Er	PNL	III (67)	SNK--TIVGSFGG-ANV
A.niger	PLD	III (127)	SNKSL-I-GE-GTSG-V
E.carot. Er	PL I	III (108)	FTKGVTILGTNGSSANF
E.carot. Er	PL III	III (108)	FTKGITIIGTNGSSANF
E.carot. EC	PLa	III (108)	FTKGLTILGTNGSSANF
E.carot. EC	PLb	III (108)	FTKGITIQGTNGSSANF
E.carot. Er	PNL	IV (86)	LRAESNSSNVIFQNLV
A.niger	PLD	IV (145)	LRMVSGVSNIIIQNIA
E.carot. Er	PNL	V (107)	AIKDNDDIQLYLNYGKGYWVDHC
A.niger	PLD	V (168)	YVWGGDAITLDEADL--VWIDHV
E.carot. Er	PL I	V (148)	Q-KMAMPSVL-ITRPN-VWIDHN
E.carot. Er	PL III	V (148)	A-QDGDAIRID-NTPN-VWIDHN
E.carot. EC	PLa	V (148)	A-KDGDAVRID-NSPN-VWIDHN
E.carot. EC	PLb	V (148)	A-QDGDAIRVD-NSPN-VWIDHN
E.carot. Er	PNL	VI (188)	R-LTICHNYYENIQVRAPGLMRYG
A.niger	PLD	VI (240)	DKVTFSGNYLYKTSGRAPKVQDNT
E.carot. Er	PL I	VI (224)	RDLTYHHNIYSDVNSRLP-LQRGG
E.carot. Er	PL III	VI (224)	RDLTYHHNIYDDVNARLP-LQRGG
E.carot. EC	PLa	VI (224)	RNLTHHHNIHSDVNSRLP-LQRGG
E.carot. EC	PLb	VI (224)	RNLTYHHNIVRDVNSRLP-LQRGG
E.carot. Er	PNL	VII (211)	YFHVFNNY-VNK-FQLAF
A.niger	PLD	VII (264)	YLHIYNNYWENN-SGHAF
E.carot. Er	PL I	VII (297)	WELRNNITSPSDF--AK
E.carot. Er	PL III	VII (297)	WELRNNVMSPADF--AK
E.carot. EC	PLa	VII (297)	WELRNNITSPSDF--AK
E.carot. EC	PLb	VII (297)	WELRNNITKPADF--SK
E.carot. Er	PNL	VIII (237)	ERNVFGSGAEKKGM--VDDKGN
E.carot. Er	PL I	VIII (187)	ESAVD---I-KKGATNVTVSYN
E.carot. Er	PL III	VIII (187)	ESAID---I-KKASTNVTISYN
E.carot. EC	PLa	VIII (187)	ESAVD---I-KKGSTNVTVSYN
E.carot. EC	PLb	VIII (187)	ESAVD---I-KKGSTNVTVSYN
E.carot. SCRI193	PLb	VIII (379)	PR--DGYYG-KKG-T-V-IKP-

[a] The deduced amino acid sequence of the *E.carotovora* Er(E.carot. Er) PNL, of the *A.niger* pectin lyase D (PLD), of the *E.carotovora* Er pectate lyases PL I and III, of the *E.carotovora* EC (E.carot. EC)pectate lyases PL a and b, and of *E.carotovora* SCRI193 (E.carot.SCRI193) pectate lyase PLb are compared.

[b] Regions (I-VIII) with putative amino acid similarities are indicated, starting with the comparison from the N-terminal of the protein. The position of the first amino acid residue of a region in each compared sequence is indicated between brackets.

[c] Within region I-VIII, identical residues at corresponding positions in the compared sequences are boxed.

the similar motif LR--S--SNVIFQNLV occurred with similar spacing (amino acid at the position 86-101 in PNL, and 145-160 in PLD) in the case of region IV among the PNL and PLD.

The nucleotide sequence of the 5'- and 3'-flanking regions. A presumed ribosome binding site (AAGGA) was found at -7 to -11 from the initiation codon of the *pnl*, which may correspond to the Shine-Dalgarno sequence (16). The sequences TTTTAT and TATTGAA were found at -131 to -136 bp and at -155 to -161 bp from the initiation codon of the *pnl* gene, respectively. The former sequence shares 4 out of 6 bases homology to the consensus sequence TATAAT in the Pribnow box (17), while the latter contains a TTG sequence and exhibits exactly the same homology with the consensus sequence of the -35 site of promoter sequence in *pelB* and *pelC* (18,19). The sequence also shows 5/7 homology with the consensus TGTTGAC at the RNA polymerase recognition site (17). The sequences presented here also resemble to those of *pel* I and III (10, 11). We previously observed that *pnl* gene was not expressed in *E.coli* JM109 harboring pTN2159, suggesting the probable presence of Lex A binding site near the promoter site of the *pnl* gene. However, whithin 290 bp upstream of the *pnl* translational start site, no sequence resembling a Lex A binding site (SOS function) was found, although there were two perfect 9 bp and 12 bp palindromic sequences present within 151 and 170 bp, and between 212 and 243 bp from *Stu*I site, respectively. The former existed in the -10 site of promoter which might be related to regulation of *pnl* transcription by binding of an unknown protein to it. The significance of these palindromic sequences in the regulation of *pnl* transcription is currently being tested. No inverted repeat which is the characteristic of transcription termination sequence was found between the stop codon of the *pnl* gene and the last codon sequenced.

Acknowledgments: We are grateful to Dr.E.Nakano and Mr.Y.Harada of Kikkoman Corporation for their kind help for the DNA sequence. We are also expressed our acknowlegment to Mr.A.Rahman in our laboratory for his critical reading of the manuscript.

REFERENCES

1. Tomizawa, S. and Takahashi, H. (1971) Agric. Biol. Chem. 35, 191-200.
2. Itoh, Y., Izaki, K. and Takahashi, H. (1980) Agric. Biol. Chem. 44, 1135-1140.
3. Zink, R. T., Engwall, J. K., McEvoy, J. L. and Chatterjee, A. K. (1985) J. Bacteriol. 164, 390-396.
4. Nishida, T., Suzuki, T., Ito, K., Kamio, Y. and Izaki, K.(1990) Biochem. Biophys. Res. Commun. 168, 801-808.
5. Henikoff, S. (1984) Gene 28, 351-359.
6. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
7. Kamimiya, S., Nishiya, T., Izaki. K. and Takahashi. H. (1974) Agric. Biol. Chem. 38, 1071-1078.
8. Kyte, J. and Doolittle, F. R. (1982) J. Mol. Biol. 157, 105-132.
9. Capaldi, A. R. and Vanderkooi, G. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 930-932.
10. Ito, K., Kobayashi, R., Nikaido, N. and Izaki, K.(1988) Agric. Biol. Chem. 52, 479-487.
11. Yoshida, A., Izuta, M., Ito, K., Kamio, Y. and Izaki, K. (1991) Agric. Biol. Chem. 55, in press.
12. Lei, S. -P., Lin, H. -C., Wang, S. -S. and Wilcox, G. (1988) Gene 62, 159-164.
13. Lei, S. -P., Lin, H. -C., Wang, S. -S., Callaway, J. and Wilcox, G. (1987) J.Bacteriol. (1987) 169, 4379-4383.
14. Hinton, J. C. D., Sidebotham, J. M., Gill, D. R. and Salmond, G. P. C. (1989) Mol. Microbiol. 3, 1785-1795.
15. Gysler, C., Harmsen, J. A. M., Kester, H. C. M., Visser, J. and Heim, J. (1990) Gene 89, 101-108.
16. Shine, J. and Dalgarno, L. (1975) Nature 254, 34-38.
17. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genetics 13, 319-353.
18. Tamaki, S. J., Gold, S., Robeson, M., Manulis, S. and Keen, N. T. (1988) J. Bacteriol. 170, 3468-3478.
19. Keen, N. T. and Tamaki, S. J. (1986) J. Bacteriol. 168, 595-606.