Cloning and Sequence Analysis of cDNA Encoding Endopolygalacturonase I from *Streptomyces purpureus*

Kazuo MIYAIRI, Mineo SENDA, Masaki WATANABE, Yuji HASUI, and Toshikatsu OKUNO

Department of Bioresearches Science, Faculty of Agriculture, Hirosaki University, and *Hirosaki University Gene Research Center, Hirosaki-shi, Aomori 036, Japan

Received October 7, 1996

Endopolygalacturonase (endoPG) I was obtained from *Streptomyces purpureus* by an improved easier purification procedure. It was found that EndoPG I consisted of three glycosylated proteins with the same isoelectric point and different molecular masses, 42, 45, and 48 kDa, respectively. However, the enzymatic deglycosylation product of endoPG I gave a single band at the position corresponding to 39 kDa on SDS-PAGE. Furthermore, the N-terminal amino acid sequences of three endoPGs were identical one another up to 20 residues. A cDNA library was constructed and positive cDNA clones encoding endoPG I were isolated by using antibody raised against the purified endoPG I. Nucleotide sequence analysis of the cDNA disclosed a 1212-bp open reading frame that encoded 403 amino acid residues. The N-terminal amino acid sequence (residues 1–20) of endoPG I coincided with the deduced amino acid sequence starting from the 25th residue. Therefore, the sequence of the first 24 residues represented a signal peptide and the remaining sequence, consisting of 379 residues, was the mature protein with molecular mass of 39.1 kDa. The deduced sequence of endoPG I showed 30–45% similarity in comparison with those of bacterial and fungal endoPGs, and the sequence of putative active site residues reported for the endoPGs was highly conserved in the sequence of endoPG I.

**Key words:** endopolygalacturonase; cDNA cloning; deduced amino acid sequence; *Streptomyces purpureus*

Previously we reported the isolation and properties of endopolygalacturonase (endoPG) I as a causative toxin of silver-leaf disease of apple trees produced by *Streptomyces purpureus*.1–4) EndoPG I is a potent pectic acid hydrolyzing enzyme, endopolygalacturonase [poly-(1,4-β-d-galacturonicide)glycanohydrolase EC 3.2.1.15], and a glycosylated protein. The ASP-4B isolate of *S. purpureus* used in our experiments produced four PGs (I, II, III, and IV) with different isoelectric points.2) Among these endoPGs, endoPG I showed the highest activity and a remarkable stability which the enzyme was not inactivated at all even after heating at 70°C for 30 min. Thereafter, it was found that this thermostable endoPG I is very useful for preparation of ohogalacturonides from pectic acid.5) For a large scale preparation a new easy purification procedure has been required since the yield in the previous method was low.

To better understand the thermostability and biological activity of endoPG I and to enable it to supply on a moderate scale, we tried to improve the purification method and to analyze the primary structure of endoPG I. In this paper, we describe a convenient purification method and some properties of endoPG I, and the cloning and sequencing of its cDNA.

**Materials and Methods**

**Microorganism and cultivation.** *S. purpureus* Persoon (ASP-4B) used in this study was supplied by the Aomori Apple Experiment Station and maintained by subculture on a potato deocction-dextrose agar medium. The medium for endoPG production was composed of potato deocction liquid (250 g/liter) containing 1% glucose. The culture was done under the same conditions except for the medium described previously.1)

**EndoPG activity assay and measurement of protein.** The enzyme reactions were done at pH 4.9 for an appropriate time using 0.2% acid insoluble polygalacturonic acid (AIP.A, DP=42) as a substrate and the activity was assayed by measuring the amount of released reducing sugars by the method of Miller and Avigad.7) One unit of the PG activity was defined as the amount of the enzyme that liberated 1 μmol of reducing sugars per min at 30°C.

The protein concentration was assayed by the method of Lowry et al.8) with bovine serum albumin as the standard, except for measuring absorbance at 280 nm for each fraction eluted from the column.

**Purification of endoPG I.** *S. purpureus* ASP-4B was cultivated for 19 days at 28°C by stationary culture in the potato deocction medium containing 1% glucose. The culture filtrate (3.8 liters) was diluted 3 times with distilled water and adsorbed on a CM52 (Whatman, U.S.A.) and DE52 (Whatman, U.S.A.) mixture column (5.1 × 30 cm). The adsorbed protein was eluted by 0.05 M phosphate buffer (pH 6.2) containing 0.2 M NaCl, followed by dialysis against 0.02 M phosphate buffer (pH 6.2) and then about 800 ml of the dialysate was lyophilized. The lyophilized sample was used as a starting material for the enzyme purification. The enzyme solution was put on a column (3.2 × 80 cm) of CM52 equilibrated with 0.02 M phosphate buffer (pH 6.5). After the column was washed with 350 ml of the same buffer, the enzyme was eluted with 300 ml of the same buffer containing 0.05 M NaCl.

**Electrophoretic analysis.** Isoelectric focusing (IEF) was done on a 7.5% (w/v) polyacrylamide gel containing 0.3% Ampholine (pH 3–10, Pharmacia Biotech) using a Atto AE-3235 apparatus (Atto, Japan). The gel was stained with Coomasie Brilliant Blue G-250 (CBB). Isoelectric points were estimated using a pH marker protein kit (pH 4.1–9.7, Oriental Yeast Co.) as standard proteins.

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was done on a 12.5% polyacrylamide gel by the method of Laemmli,9) except that samples were heated on a boiling water bath for 5 min to effect complete denaturation. The gel slabs were stained with CBB. The molecular mass was estimated using a calibration proteins kit (Pharmacia Biotech) as standard proteins.

**Deglycosylation of endoPG I.** The purified endoPG I (50 μg) was dissolved in 50 μl of 0.1 M acetate buffer (pH 6.0) and incubated with 0.1 unit of endo-β-N-acetylglucosaminidase (Seikagaku Co.) for 24 h at 30°C. The lyophilized reaction mixture was used as the sample for SDS-PAGE.

**N-Terminal amino acid sequence of endoPG I.** N-Terminal amino acid
sequencing was done with a gas-phase protein sequencer (PPSQ-10, Shimadzu Co.). EndoPG I separated on SDS-PAGE was transferred onto a PVDF membrane (Immobilon PPSQ, Nihon Millipore). The membrane was stained with CBB and cut into three pieces, each containing only a single band, and sequenced by the above apparatus. The product of deglycosylation reaction of endoPG I was also analyzed in the same manner.

Preparation of endoPG I antibodies. Antibodies were raised in rabbits by injections of 1 mg of endoPG I purified previously,6 dissolved in 1.5 ml Freund’s complete adjuvant, at intervals of two weeks for 5 weeks. After doing the Ouchterlony double diffusion test against the total proteins extracted from the culture filtrate of S. purpureum, the whole blood was collected. The precipitate obtained from the serum by addition of ammonium sulfate to 40% saturation was adsorbed to the immobolized endo-
PG I prepared from endoPG I and SMP activated-Cellulofine (Seikagaku Co.). The Cellulofine suspension was put into a column and the column was washed with 0.05 m phosphate buffer containing 0.1 m NaCl. The adsorbed anti-endoPG immunoglobulins were eluted with 0.2 m glycine-
HCl buffer (pH 2.3). Eluted fractions were rapidly neutralized by addition of Tris base and immediately dialyzed against PBS buffer (0.05 m phosphate buffer pH 7.4 + 0.1 m NaCl). The dialysate were lyophilized.

Isolation of poly(A)⁺ RNA. The mycelia of a 13-day-old culture were harvested by filtration through Whatman No. 1 paper. The mycelia mats were washed extensively with distilled water, and then ground to a fine powder in a mortar under liquid nitrogen. Total RNAs were extracted by the guanidinium,phenol method.11) The mycelium powders were homogenized in 50 ml of guanidinium thiocyanate homogenization buffer and the RNA solution obtained was put on an oligo(dT)-cellulose column to give poly(A)⁺RNA. Approximately 16 mg of total RNA and 176 μg of poly(A)⁺ RNA were recovered from 30 g of the wet mycelia.

In vivo translation assay. The rabbit reticuloocyte lysate assay was done following the manufacturer’s instructions (Promega Corporation). Each assay (50 μl) contained 20 μl of lysate mixture, 6 μCi (0.222 MBq) of [³⁵S]methionine, and 2 μg of poly(A)⁺ RNA. The mixtures were incubated at 30°C for 1 h. The translation product was immunoprecipitated with the antibody against the endoPG I, adsorbed to protein A-Cellulofine (Seikagaku Co.), and eluted with 2 x concentrated SDS-PAGE sample buffer (20 mm Tris-HCl, 2% SDS, 40% glycerol, and 0.1% bromophenol blue). The resulting translation product was analyzed by electrophoresis on SDS-PAGE (10%) and detected by fluorography using Fuji RX film. Combustheke (Boehringer Mannheim Biochemica) was used as marker proteins for molecular mass.

cDNA library construction and immunological screening. From 4 μg of poly(A)⁺ RNA, double-stranded cDNA was synthesized using a cDNA synthesis kit (Amersham Life Science). Then, the cDNA was ligated into λgt11 and packed using an in vitro packaging kit (Amersham Life Science). The constructed cDNA library, consisting of approximately 1.0 x 10⁷ independent plaques, was screened with endoPG I antibody. The inserts of the immunopositive clones were isolated and their molecular sizes checked. The cDNA insert from a selected clone was isolated and purified by agarose gel electrophoresis.

Northern blot analysis. Twenty μg of total RNA and 5 μg of the poly(A)⁺ RNA were separated by electrophoresis in 1% agarose gels and transferred onto nylon filter membranes. The filters were hybridized with ³²P-labeled endoPG I cDNA synthesized by a random labeling kit (Takara Shuzo Co.) at 65°C overnight in 0.5 m sodium phosphate, pH 7.5, 7.5% formamide, 7% SDS, 2 mm EDTA, and 100 μg/ml salmon sperm DNA. The filter was washed three times at 65°C with 2 x SSC (0.3 m sodium citrate, 0.3 m NaCl, pH 7.0)-0.1% SDS or 37°C with 2 x SSC-0.1% SDS.

cDNA sequence analysis. The purified cDNA insert was subcloned into pBluescript KS⁺ (Stratagene). The construct was double-digested with XbaI and XhoI and deletion clones were prepared by digestion of cDNA fragments with exonuclease III and mung bean nuclease (Takara Shuzo Co.). DNA was sequenced by an ALFred DNA Sequencer II (Pharmacia Biotech) using Auto Read sequencing Kit (Pharmacia Biotech). Nucleotide sequences were analyzed using the DNAMAS-Mac software programs (Hitachi Software Engineering Co.) for prediction of amino acid sequences. A protein sequence similarity search was done by the method of Lipman and Pearson12) using the SWISS-PROT database.

Results and Discussion

Purification of endoPG I from S. purpureum

An improved method for purification of endoPG I was required to supply a large amount of endoPG I as an oligogalacturonic acid preparing enzyme. The purification procedure of endoPG I is summarized in Table. EndoPG I was purified by two steps in 37% yield. The results of CM52 column chromatography are shown in Fig. 1. The eluted protein peaks 1, 2, and 3 were estimated to correspond to endoPG III and IV, and I, respectively.21) The fractions corresponding to peak 1 were pooled as a purified endoPG I. The use of the CM52 and DE52 mixture column gave a good yield purification, time saving, and recovery of all the isozymes of endoPGs in the culture filtrate.

Molecular masses and N-terminal amino acid sequences of endoPG I

The purified endoPG I showed a single band on IEF-
PAGE (Fig. 2A) and its isoelectric point was estimated to be pH 8.8, somewhat higher than the value (8.5) by preparative electrophocusing reported previously.4) On the other hand, the SDS-PAGE indicated three bands (Ia, Ib, and Ic) as shown in Fig. 2B and their MWs 42, 45, and 48 kDa, respectively. These three bands, however, changed into a sharp single band of MW 39.5 kDa (Fig. 2B, lane 3) upon deglycosylation with endo-β-N-acetylglucosaminidase. Furthermore, all the three bands showed endoPG activity on native PAGE (pH 4.0) by the activity staining method13) (data not shown). It seems that the observation in this time differed from the previous one is attributable

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Purification of EndoPG I from S. purpureum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step</td>
<td>Total activity (units)</td>
</tr>
<tr>
<td>Culture filtrate</td>
<td>104,800</td>
</tr>
<tr>
<td>CM52, DE52</td>
<td>90,100</td>
</tr>
<tr>
<td>CM52</td>
<td>38,700</td>
</tr>
</tbody>
</table>

Fig. 1. Purification of EndoPG I by CM52 Column Chromatography.}

The chromatography was done under the conditions described in Materials and Methods. The position of the arrow indicates change of elution buffer from 0.02 m phosphate buffer (pH 6.5) to the same buffer containing 0.05 m NaCl. Fractions (5 ml) were collected and assayed for OD₅₈₀ and endoPG activity (●).
to a variation of the enzyme productivity of *S. purpureum* (ASP-4B). The previously reported endoPG I with 41 kDa of MW on SDS-PAGE\(^3\) may correspond to endoPG Ia, from comparison of the MWs.

The N-terminal amino acid sequence of the three bands of endoPG I were identical and found to be ATXTVKSV-DDAKDIAAGXS A (X: unidentified residue). Therefore, these results suggest that the three endoPGs have the same or very similar primary structures.

**In vitro translation**

*In vitro* translations were examined using poly(A)\(^+\) RNAs from the 13-day-old mycelia. The translation protein was detected as a clear band at 39 kDa on the autoradiogram (Fig. 3A). Although the endoPG activity is highest in the broth cultured for about 20 days, the clear band at 39 kDa have not been detected in poly(A)\(^+\) RNAs from the 19-day-old mycelia (data not shown). The 39-kDa protein appeared to be an endoPG without carbohydrates. This result indicated that the poly(A)\(^+\) RNA preparation indeed contained mRNA encoding endoPG I.

**Isolation of endoPG I cDNA clone**

Among approximately 10,000 recombinant plaques 5 positive clones were isolated by screening with endoPG I antibody. One immunopositive clone was purified and the size of cDNA insert isolated after *EcoRI* digestion was about 1.4 kilobase pairs (kb).

**Northern blot analysis of RNA of *S. purpureum***

Northern hybridization using endoPG I cDNA labeled by \(^{32}\)P showed a single band on a fluorogram (Fig. 3B). The size of hybridized poly(A)\(^+\) RNA was estimated to be approximately 1.5 kb, therefore this cDNA used as a probe seemed to be the full length needed to encode the amino acid sequence of endoPG I.
poly(A)$^+$ tail.

The N-terminal amino acid sequence of endoPG I identified by a protein sequencer was found at positions 25 to 44 in the deduced amino acid sequence, following the putative signal peptide consisting of 24 amino acid residues. The predicted mature endoPG consisted of 379 amino acids with a calculated molecular mass of 39,119 Da. This data coincided well with the molecular mass of the deglycosylated endoPG I of 39 kDa on SDS-PAGE. Also, the amino acid composition of endoPG I$^4$ from the amino acid analysis, except for the proline content, agreed well with that from the deduced amino acid sequence. These results suggested that the cDNA insert contained the coding region of the mature endoPG I protein.

Furthermore, the deduced amino acid sequence showed that there were six possible N-glycosylation sites, namely Asn 116, 185, 303, 326, 332, and 398. Therefore, it is considered that the difference in molecular mass among the three kinds of molecules of endoPG I depends on the number of sugar chains in molecular and/or different structures of the sugar chains.

**Comparison of the amino acid sequence of endoPG I with those of other PGs**

Comparison of the amino acid sequence of endoPG I with those of other endopolygalacturonases showed that endoPG I had 30–45% homology with endoPGs from Sclerotinia sclerotiorum, Fusarium moniliforme, Coeliobolus carbonum, Aspergillus oryzae, A. niger, Erwinia carotovora, Pseudomonas solanacearum, avocado, and tomato. However, endoPG I showed the low sequence similarity of 29.9% with the exoPG from Arabidopsis thaliana (F. R. Quigley, unpublished databases, 1993).

In a comparison of the sequences, endoPG I shows fairly high similarity with fungal endoPGs from F. solani$^4$ (44.8%) and C. carbonum$^1$ (43.6%), but lower similarity, 30%–32%, with bacterial endoPGs. No similarity was found between the endoPG I and other pectin-degrading enzymes other than PG.

The amino acid sequence alignments of the putative conserved segments of endoPGs$^{17,23}$ are shown in Fig. 5. The amino acid sequences of conserved regions common in endoPGs from various sources are also found in endoPG I. The existence of histidine and aspartic acid residues in the active site are considered to be essential for the catalytic action of endoPGs,$^{24}$ hence, their residues in the conserved regions shown in Fig. 5 may be the corresponding amino acid residues in the active site.

In this study we could not find a reason why endoPG I

---

**Fig. 4.** Nucleotide and Deduced Amino Acid Sequence of the cDNA Encoding EndoPG I.
The underline below the amino acid sequence and the double underline below the nucleotide sequence indicate the signal peptide and the putative polyadenylation signal, respectively. The wave underline shows the N-terminal amino acid sequence analyzed directly from endoPG I. The putative conserved regions are boxed. The Gothic letter of N indicates a possible site for an N-linked saccharide chain. The nucleotide sequence has been deposited in the DDBJ, DDBJ, EMBL, and NCBI nucleotide sequence databases with the accession No. D 45072.
has a higher thermostability than other fungal endoPGs, however, much information was obtained to resolve the three-dimensional structure including the sugar chain linked positions.

Acknowledgment. We thank Dr. H. Matsue, Aomori Advanced Industrial Technology Center, for analysis of N-terminal amino acid sequences.

References